PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C07H 21/04 — A1

(11) International Publication Number: **WO 96/33207**

(43) International Publication Date: 24 October 1996 (24.10.96)

(21) International Application Number: PCT/US96/05480

(22) International Filing Date: 18 April 1996 (18.04.96)

(30) Priority Data:
08/425,684 18 April 1995 (18.04.95) US

(60) Parent Application or Grant
(63) Related by Continuation
US 08/425,684 (CIP)
Filed on 18 April 1995 (18.04.95)

(71) Applicant *(for all designated States except US)*: GLAXO GROUP LIMITED [GB/GB]; Glaxo Wellcome House, Berkeley Avenue, Greenford, Middlesex UB6 0NN (GB).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: STEMMER, Willem, P., C. [NL/US]; 108 Kathy Court, Los Gatos, CA 95030 (US). LIPSHUTZ, Robert, J. [US/US]; 970 Palo Alto Avenue, Palo Alto, CA 94301 (US).

(74) Agents: DUNN, Tracy, J. et al.; Townsend and Townsend and Crew, Steuart Street Tower, 20th floor, One Market Plaza, San Francisco, CA 94105 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: END-COMPLEMENTARY POLYMERASE REACTION

(57) Abstract

The present invention is directed to a process for amplifying and detecting any target nucleic acid sequence contained in a nucleic acid or mixture thereof and for assembling large polynucleotides from component polynucleotides, each involving generating concatemers formed by PCR amplification of overlapping fragments.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgystan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic			SE	Sweden
CG	Congo	KR	Republic of Korea	SG	Singapore
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LR	Liberia	SZ	Swaziland
CS	Czechoslovakia	LT	Lithuania	TD	Chad
CZ	Czech Republic	LU	Luxembourg	TG	Togo
DE	Germany	LV	Latvia	TJ	Tajikistan
DK	Denmark	MC	Monaco	TT	Trinidad and Tobago
EE	Estonia	MD	Republic of Moldova	UA	Ukraine
ES	Spain	MG	Madagascar	UG	Uganda
FI	Finland	ML	Mali	US	United States of America
FR	France	MN	Mongolia	UZ	Uzbekistan
GA	Gabon	MR	Mauritania	VN	Viet Nam

END-COMPLEMENTARY POLYMERASE REACTION

TECHNICAL FIELD

The present invention relates generally to the field of recombinant DNA technology and, more particularly, to improved methods for producing amplified heterogeneous populations of polynucleotides from limited quantities of DNA or RNA or other nucleic acids. The invention provides compositions and methods for a chain reaction amplification of a target polynucleotide species using a thermostable polymerase or other suitable polynucleotide polymerase compatible with the method.

BACKGROUND

Selective amplification of polynucleotides represents a major research goal of molecular biology, with particular importance in diagnostic and forensic applications, as well as for general manipulations of genetic materials and laboratory reagents.

The polymerase chain reaction (PCR) is a method by which a specific polynucleotide sequence can be amplified in vitro. PCR is an extremely powerful technique for amplifying specific polynucleotide sequences, including genomic DNA, single-stranded cDNA, and mRNA among others. As described in U.S. Patent Nos. 4,683,202, 4,683,195, and 4,800,159 (which are incorporated herein by reference), PCR typically comprises treating separate complementary strands of a target nucleic acid with two oligonucleotide primers to form complementary primer extension products on both strands that act as templates for synthesizing copies of the desired nucleic acid sequences. By repeating the separation and synthesis steps in an automated system, essentially exponential duplication of the target sequences can be achieved.

A number of variations of the basic PCR methodology have been described. U.S. Patent No. 5,066,584 discloses a method wherein single stranded DNA can be generated by the polymerase chain reaction using two oligonucleotide primers,

one present in a limiting concentration. U.S. Patent No. 5,340,728 discloses an improved method for performing a nested polymerase chain reaction (PCR) amplification of a targeted piece of DNA, wherein by controlling the annealing times and concentration of both the outer and the inner set of primers according to the method disclosed, highly specific and efficient amplification of a targeted piece of DNA can be achieved without depletion or removal of the outer primers from the reaction mixture vessel. U.S. Patent No. 5,286,632 discloses recombination PCR (RPCR) wherein PCR is used with at least two primer species to add double-stranded homologous ends to DNA such that the homologous ends undergo in vivo recombination following transfection of host cells.

Horton et al. (1989) Gene 77: 61, discloses a method for making chimeric genes using PCR to generate overlapping homologous regions. In the Horton method, fragments of different genes that are to form the chimeric gene are generated in separate polymerase chain reactions. The primers used in these separate reactions are designed so that the ends of the different products of the separate reactions contain complementary sequences. When these separately produced PCR products are mixed, denatured and reannealed, the strands having matching sequences at their 3'-ends overlap and act as primers for each other. Extension of this overlap by DNA polymerase produces a molecule in which the original sequences are spliced together to form the chimeric gene.

Silver and Keerikatte (1989) J. Virol. 63: 1924 describe another variation of the standard PCR approach (which requires oligonucleotide primers complementary to both ends of the segment to be amplified) to allow amplification of DNA flanked on only one side by a region of known DNA sequence. This technique requires the presence of a known restriction site within the known DNA sequence and a similar site within the unknown flanking DNA sequence which is to be amplified. After restriction and recircularization, the recircularized fragment is restricted at an unique site between the two primers and the resulting linearized fragment is used as a template for PCR amplification.

Triglia et al. (1988) Nucl.Acids Res. 16: 8186, describe an approach which requires the inversion of the s quence of interest by circularization and re-opening at a site distinct from the one of interest, and is called "inverted PCR." A fragment is first created in which two unknown sequences flank on either side a region of known DNA sequence. The fragment is then circularized and cleaved with an unique restriction endonuclease which only cuts within the known DNA sequence creating a new fragment containing all of the DNA of the original fragment but which is then inverted with regions of known sequence flanking the region of unknown sequence. This fragment is then utilized as a PCR substrate to amplify the unknown sequence.

Vallette et al. (1989) Nucl.Acids Res. 17: 723 disclose using PCR in a specific approach which involves using a supercoiled plasmid DNA as a template for PCR and a primer bearing a mutated sequence which is incorporated into the amplified product. Using this method, DNA sequences may be inserted only at the 5'-end of the DNA molecule which one wishes to alter. Mole et al. (1989) Nucl.Acids Res. 17: 3319, used PCR to create deletions within existing expression plasmids. However, PCR was performed around the entire plasmid (containing the fragment to be deleted) from primers whose 5'-ends defined the region to be deleted. Self-ligation of the PCR product recircularized the plasmid.

U.S. Patent No. 5,279,952 discloses a method for using PCR to generate mutations (e.g., deletions) and chimeric genes by forming head-to-tail concatemers of a known starting sequence and employing at least two PCR primers to amplify a DNA segment which is altered as compared to the known starting sequence.

Jones and Howard (1990) BioTechniques 8: 178, report a site-specific mutagenesis method using PCR, termed recombinant circle PCR (RCPCR). In RCPCR, separate PCR amplifications (typically two) of a known polynucleotide generate products that, when combined, denatured, and annealed, form double-stranded DNA with discrete, cohesive

single-stranded ends designed so that they may anneal and form circles of DNA.

Oliner et al. (1993) Nucl. Acids. Res. 21: 5192, report a method for engineering PCR products to contain terminal sequences identical to sequences at the two ends of a linearized vector such that co-transfection of the PCR product and linearized vector into a recombination-competent host cell results in formation of a covalently linked vector containing the PCR product, thus avoiding the need for in vitro ligation.

In spite of such recent advances, including PCR and its various modifications noted above, there exists a need for improved methods of identifying and cloning polynucleotides, for accurate in vitro amplification of selected polynucleotides, and for facile assembly of polynucleotides from a mixture of component oligonucleotides or polynucleotides without necessitating the use of DNA ligase. In particular, there is a need for a PCR amplification method which can be performed with (1) only a single primer species, or (2) with multiple overlapping polynucleotide fragments (or oligonucleotides) in the absence of a conventional PCR primer, and which can result in formation of an amplified product which can be a concatemer and/or which can be a covalently-closed circle. The present invention fulfills these and other needs.

The references discussed herein are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that the inventors are not entitled to antedate such disclosure by virtue of prior invention. All publications cited are incorporated herein by reference.

SUMMARY OF THE INVENTION

A basis of the present invention is the use of polymerase-mediated chain extension, such as for example PCR, in combination with at least two polynucleotides having complementary ends which can anneal whereby at least one of said polynucleotides has a free 3'-hydroxyl capable of polynucleotide chain elongation by a DNA polymerase, such as a

thermostable polymerase (e.g., Thermus aquaticus (Taq) polymerase or Thermococcus litoralis (Vent™, New England Biolabs) polymerase or TthI polymerase (Perkin-Elm r). Although the method can be practiced using PCR, in some embodiments either a single primer species or no primer whatsoever is required, and hence PCR is not a necessary component of the general method.

In one embodiment, a target polynucleotide is contacted with a "bivalent primer" typically comprising an oligonucleotide having two regions of complementarity to the target polynucleotide: (1) a first portion which is in the 5' portion of the primer and which is substantially complementary to a sequence in the 3' portion of the sequence to be amplified (target sequence) in the target polynucleotide, and (2) a second portion which is in the 3' portion of the primer and which is substantially complementary to a sequence in the 5' portion of the sequence to be amplified (target sequence) in the target polynucleotide. The contacting is performed under conditions suitable for hybridization of the bivalent primer to the target polynucleotide, most often following thermal denaturation of the target polynucleotide if it is initially present as a double-stranded form. The target polynucleotide may be substantially homogeneous or may be present in a mixture of polynucleotide species (e.g., in a genome, biological sample, or mixture of synthetic polynucleotides). Subsequent or concomitant with the contacting of the bivalent primer to the target polynucleotide, a polynucleotide polymerase, such as a thermostable DNA polymerase, catalyzes, under suitable reaction conditions, polynucleotide synthesis (chain elongation) primed from the 3'-hydroxyl of the annealed bivalent primer to form a strand complementary to the target sequence, thereby forming a nascent complementary strand. Following completion of the nascent complementary strand spanning the target sequence, the target polynucleotide and the nascent strand are denatured, typically by elevation of temperature, and allowed to reanneal, typically by reduction of temperature, with another molecule of the bivalent primer

species or with a complem ntary strand of a target polynucleotide or an amplified copy thereof. The denatured nascent strand species following th first longation cycle will contain a copy of the target sequence and has a terminal repeat of its 5'-terminal sequence at the 3' terminus, resulting from the bivalent primer, and wherein the terminal repeat is of sufficient length to support annealing under PCR conditions to an overlapping complementary strand in a head-to-tail arrangement (see, Fig. 1). Following reannealing, the described polymerase elongation/denaturation/reannealing cycle is repeated from 1 to about 100 times as desired, resulting in formation of amplified product which comprises head-to-tail concatemers of the target sequence. The concatemers typically increase in length as the number of amplification cycles increase and as the amount of bivalent primer decreases. Following amplification forming concatameric head-to-tail repeats of the target sequence, the concatemer(s) can optionally be resolved, such as (1) by cleaving with a restriction endonuclease which cuts within (or at the termini of) the concatemeric unit(s), (2) by homologous recombination between concatemer units to form covalently closed circles, or (3) by cleavage with a restriction endonuclease followed by ligation with DNA ligase to form covalently closed circles and/or by direct transformation into host cells for in vivo ligation.

Often, a target polynucleotide sequence which is amplified as described above will form amplification intermediates in the form of cyclized DNA or spiral DNA (see, Fig. 2), as a result of the 3' terminus of an overlapped nascent strand annealing to the 3' terminus of an overlapped complementary strand forming a cyclized structure similar to a gapped circle; the cyclized structure has a strand with an extendable 3'-hydroxyl which can be extended with a DNA polymerase substantially lacking exonuclease activity (e.g., a thermostable polymerase such as Vent(exo$^-$)^{TM} or Klenow fragment, etc.) in a rolling circle format whereby the leading terminus of the nascent strand continually displaces the lagging portion of the nascent strand (see, Fig. 2) producing

a concatemeric single strand emanating from th rolling circle intermediate. Most oft n, such rolling circle intermediates will form under dilute conditions more favorable to intramolecular cyclization of overlapped strands. Once a rolling circle intermediate is established, the template need not be denatured in order to continue amplification of the target sequence as in conventional PCR, thus avoiding the necessity of multiple thermal cycles of PCR to denature template (and the resultant time loss needed for heating and cooling). Often, however, the template is repeatedly denatured, annealed, and extended with polymerase in the presence of ribonucleotide or deoxyribonucleotides under suitable reaction conditions.

Furthermore, whether the method generates a rolling circle intermediate or linear concatemers, an advantage of the method is that it requires a substantially reduced amount of primer (bivalent primer) as compared to conventional PCR, since following the initial cycle(s) an increasing percentage of the priming of nascent strand synthesis is primed from 3'-hydroxyl groups of the amplified strands, rather than from the oligonucleotide primer(s). In the case of a rolling circle intermediate, theoretically only a single bivalent primer molecule is necessary to generate the rolling circle which then can produce multiple concatenated copies by rolling circle-style polymerase catalysis using a polymerase capable of strand displacement of the lagging edge of the nascent strand as replication proceeds around the cyclized template.

In an embodiment, a product polynucleotide is assembled from a plurality of component polynucleotides by formation of overlapped strands of alternating polarity and having substantially complementary termini (see, Fig. 3). This method employs a series of overlapping substantially complementary termini to determine the linear order of component sequences in the final product. Concomitant with or subsequent to formation of the overlapped strands of the component polynucleotides in a reaction, a polynucleotide polymerase (e.g., a thermostable DNA polymerase) under suitable reaction conditions catalyzes strand elongation from

the 3'-hydroxyl portions of the overlapped (annealed) joints, filling in the portion between joints and processively displacing or processively degrading exonucleolytically the 5' termini of downstream component strands of the same polarity as the nascent strand elongates. After a cycle of chain elongation forming substantially double-stranded polynucleotides, the reaction conditions are altered (typically by increasing the temperature) to effect denaturation of the double-stranded polynucleotides, followed by altering the reaction conditions to permit reannealing of complementary strands or portions thereof (i.e., overlapping termini) to form molecules having overlapped termini (joints), and chain elongation by a polynucleotide polymerase under suitable reaction conditions catalyzes strand elongation from the 3'-hydroxyl portions of the overlapped (annealed) joints, as in the first cycle. One to about 100 cycles of denaturation/annealing/polymerization can be performed to generate a product comprising the component polynucleotide sequences covalently linked in linear order according to the order of the overlapping joints. In this embodiment, a product polynucleotide can be constructed from a plurality of smaller component polynucleotides (typically oligonucleotides) and enables assembly of a variety of products with alternate substitutable polynucleotide components at a given position serving as structural "alleles" (see, Fig. 4). The component polynucleotides are often provided in single-strand form, but may initially be present in double-strand form and be denatured (typically by elevated temperature) for the assembly of the product by PCR amplification. Substantially any type of product polynucleotide can be assembled in this way, including cloning and expression vectors, viral genomes, gene therapy vectors, genes (including chimeric genes), polynucleotides encoding peptide libraries, protein libraries, vector libraries, viral libraries, and the like. In a variation, one or more of the component polynucleotides represents a site-directed mutation or variable-sequence kernel. In a variation, PCR employing a low-fidelity polymerase is used to introduce additional sequence variation

into the product polynucleotide(s) during amplification cycles. The method can be used to produce a library of sequence-variant product polynucleotides, if desired.

In an embodiment of the invention, very long distance PCR is provided, wherein PCR or other suitable amplification method is used to generate, in a single reaction or in parallel reactions which are subsequently pooled, a set of overlapping large DNA fragments which can be denatured and annealed to form very large (e.g., greater than 25 to 50 kilobases) DNA structures composed of overlapped single strands of DNA having alternating polarity with each overlapped joint providing an extendable 3'-hydroxyl group for forming phosphodiester bonds catalyzed by a polynucleotide polymerase in the presence of free ribonucleotide or deoxyribonucleotides. Typically, the method comprises forming at least three overlapping polynucleotides, wherein the 3' terminus of a first single-stranded polynucleotide is substantially complementary to the 3' terminus of a second single-stranded polynucleotide of the opposite polarity, and wherein the 5' terminus of said second single-stranded polynucleotide is substantially complementary to the 3' terminus of a third single-stranded polynucleotide having polarity identical to said first single-stranded polynucleotide, thereby generating an overlapped structure capable of chain elongation by a suitable polymerase to generate a double-stranded product spanning the three initial overlapped polynucleotides. With such a method, polynucleotides of 50 kb to 100 kb or more can be generated by a facile amplification method capable of generating amplification products much longer than is possible with conventional long-range PCR methods. The method can comprise parallel processing PCR reactions, wherein a plurality of primer sets are employed in a single reaction or multiple reactions which are subsequently pooled, each primer sets priming the PCR amplification of a polynucleotide sequence which comprises terminal sequences which are complementary to terminal sequences in at least on other amplification product produced by a different primer set, thus generating a set of

overlapping PCR products with which a large product spanning the entire set of PCR products is generated by end-complementary polymerase reaction.

In some embodiments of the invention, the polynucleotides product(s) generated thereby are labelled, such as with radioisotopic, biotinyl, or fluorescent label moieties, by incorporation of labelled ribonucleotide or deoxyribonucleotides or the like into nascent polynucleotide by polymerase-mediated catalysis.

The invention also provides kits comprising a bivalent primer polynucleotide and/or a plurality of component polynucleotides and instructions for use describing the present end-complementary amplification method disclosed herein. Frequently, a polynucleotide polymerase, such as a thermostable DNA polymerase (Taq or Vent™ polymerase) is also present in the kit. Optionally, one or more target polynucleotides may be provided in the kit, such as for calibration and/or for use as a positive control to verify correct performance of the kit.

In an embodiment, the invention provides a method termed continuous multiplex amplification which affords amplification of a plurality of initially unlinked polynucleotide species at substantially comparable amplification rates by forming a linked amplification product wherein the plurality of initially unlinked polynucleotide sequences are linked by end-complementary amplification. An amplification unit, termed an amplicon, comprising at least one copy of each member of the plurality of initially unlinked polynucleotide species is formed by one or more cycles of end-complementary amplification. From one to about 100, typically three to 35, amplification cycles can be conducted and result in formation of a population of linked amplification products, which can comprise concatemers of said amplicon. The amplification products can be linear or circular, as desired, based on appropriate selection of the bivalent primers. In a variation, the amplification product is cleaved with a nucleolytic agent, such as a restriction enzyme which cuts at least one restriction site present in the amplicon, DNase,

nuclease S1, bleomycin, ionizing radiation, or the like or by other suitable cleaving means.

A further understanding of the nature and advantages of the invention will become apparent by reference to the remaining portions of the specification and drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1. Schematic of bivalent primer and concatemer formation in end-complementary PCR.

Fig. 2 Schematic depiction of cyclized intermediates and rolling circle amplification.

Fig. 3 Schematic of overlapping fragment PCR for construction and amplification of larger products from component polynucleotides.

Fig. 4 Schematic of multiple "alleles" with overlapping fragment PCR.

Fig. 5A-E shows oligonucleotides used to construct the 2.7 kb circular plasmid p182SfiI by end-complementary polymerase reaction.

Fig. 6 Schematic of plasmid construction by overlapping fragment PCR as performed in Experimental Examples.

Fig. 7 Agarose gel electrophoresis of products generated during construction of the 2.7 kb circular plasmid p182SfiI by end-complementary polymerase reaction. The circled letters refer to aliquots removed from various amplification reactions: (A) is the mixture of oligonucleotides without polymerase, (B) is the product of the first set of amplification cycles, (C) is the product of the second set of amplification cycles, (D) is the final product.

Fig. 8 Schematic for end-complementary polymerase reaction (ECPR) in conjunction with parallel-processing PCR to amplify very large polynucleotides, such as those larger than can be amplified reliably by conventional PCR using only a single primer set.

Figs. 9A-9C Schematic of continuous circular multiplex amplification methodology exemplifying amplification of two unlinked polynucleotides, double-stranded ABC/A'B'C'

and double-stranded DEF/D'E'F'. A and A', B and B', C and C', D and D', E and E', F and F' each represent a set of complementary (or substantially complementary) polynucleotide sequences present in the initially unlinked polynucleotides. X and X', Y and Y' each represent a set of complementary (or substantially complementary) polynucleotide sequences present in the bivalent primers CYD, C'X'D', FYA, and F'X'A'. In some embodiments, X and X' and/or Y and Y' can be omitted. Fig. 9A shows the initial (starting) conditions with a plurality of polynucleotide species (shown for example as two double-stranded molecules) to be amplified by continuous circular multiplex amplification. The second step, "Anneal with Primers", shows the structural features of the bivalent primers and their mode of annealing to single-stranded polynucleotide species, shown as denatured double-stranded complementary polynucleotides. The third step, "Extend and Melt", shows the amplification products after a cycle of extension primed by the bivalent primers; the amplification products are denatured for a subsequent round(s) of amplification. Fig. 9B shows the possible modes of reannealing of the amplification products generated from the first round of amplification. Fig. 9C shows the amplification products which can result from the second round (or subsequent rounds) of amplification; each of the product molecules is capable of self-replication as they have complementary ends, and they can also cross-replicate. Each of the product molecules comprises copies of the initially unlinked polynucleotide sequences in equimolar ratios.

Figs. 10A-10C Schematic of continuous linear multiplex amplification methodology exemplifying amplification of two unlinked polynucleotides, double-stranded ABC/A'B'C' and double-stranded DEF/D'E'F'. A and A', B and B', C and C', D and D', E and E', F and F' each represent a set of complementary (or substantially complementary) polynucleotide sequences present in the initially unlinked polynucleotides. X and X', Y and Y' each represent a set of complementary (or substantially complementary) polynucleotide sequences. X' and Y are present in the bivalent primers CYD and C'X'D'. The

univalent primers are F' and A. In some embodiments, X and X' and/or Y and Y' can be omitted. Fig. 10A shows the initial (starting) conditions with a plurality of polynucleotide species (shown for example as two double-stranded molecules) to be amplified by continuous circular multiplex amplification. The second step, "Anneal with Primers", shows the structural features of the bivalent and univalent primers and their mode of annealing to single-stranded polynucleotide species, shown as denatured double-stranded complementary polynucleotides. The third step, "Extend and Melt", shows the amplification products after a cycle of extension primed by the bivalent and univalent primers; the amplification products are denatured for a subsequent round(s) of amplification. Fig. 10B shows the possible modes of reannealing of the amplification products generated from the first round of amplification. Fig. 10C shows the amplification products which can result from the second round (or subsequent rounds) of amplification; each of the product molecules comprises copies of the initially unlinked polynucleotide sequences in equimolar ratios.

Figs. 11A-11C Schematic of continuous circular multiplex amplification methodology exemplifying amplification of two possibly unlinked polynucleotides embedded in distinct locations in a genome or pool of DNA molecules. Fig. 11A shows the initial (starting) conditions with a plurality of polynucleotide sequences (shown for example as two double-stranded sequences embedded in discrete genomic locations) to be amplified by continuous circular multiplex amplification. First, the genomic sequences are amplified using a low concentration of conventional amplification primers (shown as PCR primers; C',F',A, and D) as indicated under "Anneal #1". The concentration of conventional primers, initial copy number, and number of amplification cycles is such that primers for rapidly extending fragments are consumed and slowly extending sequences are allowed to amplify. Fig. 11B shows that bivalent primers (FT3'XT7A and CYD) are used in subsequent rounds of amplification. In this example, one of the bivalent primers (FT3'XT7A) comprises the sequences for

one or more promoter sequence, in this case a T3 promoter and a T7 promoter oriented in opposite transcriptional polariti s. The mode of hybridization of the bivalent primers to denatured amplification product is shown under "Anneal with Primers", and the resultant amplification products are shown under "Extend and Melt". Fig. 11C shows possible modes of reannealing of the denatured products of amplification using the bivalent primers under "Reannealing". Examples of the resultant products of self-primed amplification are shown under "Extend". Each of the product polynucleotides shown have complementary ends and are capable of self-replication and cross-replication. The sequences X and Y, if present, can comprise restriction sites, if desired.

Figure 12 Arsenate, arsenite and antimony resistance for E. coli strain TG1, without a plasmid, with plasmid pGJ103 with the wild type ars operon, or with pGJ103 mutagenized by three cycles of PCR shuffling. Cells grown overnight in LB with 2 mM, 10 mM or 128 mM arsenate were diluted 10,000-times into LB with added oxyanions as indicated and turbidity was measured after 16 hours growth at 37°C. Equal amounts of cells (OD600) were plated on plates with a range of arsenate concentrations and grown overnight at 37°C. Cell growth was quantitated by resuspending the cells and measuring the OD600.

Figure 13 Cells as in Figure 12 were washed and suspended in triethanol amine buffer and exposed to 3 mM ^{73}As-arsenate. Samples were removed periodically, heated to 100°C, and centrifuged. ^{73}As-arsenate and ^{73}As-arsenite were quantitated after thin layer chromatographic separation.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred

methods and materials are described. For purposes of the present inv ntion, the following terms are defined below.

Definitions

Unless specified otherwise, the conventional notation used herein portrays polynucleotides as follows: the lefthand end of single-stranded polynucleotide sequences is the 5' end; the lefthand direction of double-stranded polynucleotide sequences is referred to as the 5' direction. The direction of 5' to 3' addition of nascent RNA transcripts is referred to as the transcription direction; sequence regions on the DNA strand having the same sequence as the RNA and which are 5' to the 5' end of the RNA transcript are referred to as "upstream sequences"; sequence regions on the DNA strand having the same sequence as the RNA and which are 3' to the 3' end of the coding RNA transcript are referred to as "downstream sequences".

As used herein, the term "polynucleotide" refers to a polymer composed of a multiplicity of nucleotide units (ribonucleotide or deoxyribonucleotide or related structural variants) linked via phosphodiester bonds. A polynucleotide can be of substantially any length, typically from about 10 nucleotides to about $1x10^9$ nucleotides or larger. As used herein, an "oligonucleotide" is defined as a polynucleotide of from 6 to 100 nucleotides in length. Thus, an oligonucleotide is a subset of polynucleotides.

The term "naturally-occurring" as used herein as applied to an object refers to the fact that an object can be found in nature. For example, a polypeptide or polynucleotide sequence that is present in an organism (including viruses) that can be isolated from a source in nature and which has not been intentionally modified by man in the laboratory is naturally-occurring. Generally, the term naturally-occurring refers to an object as present in a non-pathological (undiseased) individual, such as would be typical for the species.

The term "corresponds to" is used herein to mean that a polynucleotide sequence is homologous (i.e., is

identical, not strictly evolutionarily related) to all or a portion of a reference polynucleotide sequence. In contradistinction, th term "complementary to" is used herein to mean that the complementary sequence is homologous to all or a portion of a reference polynucleotide sequence. For illustration, the nucleotide sequence "TATAC" corresponds to a reference sequence "TATAC" and is complementary to a reference sequence "GTATA".

The following terms are used to describe the sequence relationships between two or more polynucleotides: "reference sequence", "comparison window", "sequence identity", "percentage of sequence identity", and "substantial identity". A "reference sequence" is a defined sequence used as a basis for a sequence comparison; a reference sequence may be a subset of a larger sequence, for example, as a segment of a full-length cDNA or gene sequence given in a sequence listing, or may comprise a complete cDNA or gene sequence. Generally, a reference sequence is at least 12 nucleotides in length, frequently at least 15 to 18 nucleotides in length, and often at least 25 nucleotides in length. Since two polynucleotides may each (1) comprise a sequence (i.e., a portion of the complete polynucleotide sequence) that is similar between the two polynucleotides, and (2) may further comprise a sequence that is divergent between the two polynucleotides, sequence comparisons between two (or more) polynucleotides are typically performed by comparing sequences of the two polynucleotides over a "comparison window" to identify and compare local regions of sequence similarity.

A "comparison window", as used herein, refers to a conceptual segment of at least 12 contiguous nucleotide positions wherein a polynucleotide sequence may be compared to a reference sequence of at least 12 contiguous nucleotides and wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) of 20 percent or less as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. Optimal alignment of sequences for aligning a comparison window may be conducted by

the local homology algorithm of Smith and Waterman (1981) Adv. Appl. Math. 2: 482, by the homology alignment algorithm of Needleman and Wunsch (1970) J. Mol. Biol. 48: 443, by the search for similarity method of Pearson and Lipman (1988) Proc. Natl. Acad. Sci. (U.S.A.) 85: 2444, by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package Release 7.0, Genetics Computer Group, 575 Science Dr., Madison, WI), or by inspection, and the best alignment (i.e., resulting in the highest percentage of homology over the comparison window) generated by the various methods is selected.

The term "sequence identity" means that two polynucleotide sequences are identical (i.e., on a nucleotide-by-nucleotide basis) over the window of comparison. The term "percentage of sequence identity" is calculated by comparing two optimally aligned sequences over the window of comparison, determining the number of positions at which the identical nucleic acid base (e.g., A, T, C, G, U, or I) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison (i.e., the window size), and multiplying the result by 100 to yield the percentage of sequence identity. The terms "substantial identity" as used herein denotes a characteristic of a polynucleotide sequence, wherein the polynucleotide comprises a sequence that has at least 80 percent sequence identity, preferably at least 85 percent identity and often 90 to 95 percent sequence identity, more usually at least 99 percent sequence identity as compared to a reference sequence over a comparison window of at least 20 nucleotide positions, frequently over a window of at least 25-50 nucleotides, wherein the percentage of sequence identity is calculated by comparing the reference sequence to the polynucleotide sequence which may include deletions or additions which total 20 percent or less of the reference sequence over the window of comparison. The reference sequence may be a subset of a larger sequence.

The primers herein are selected to be substantially complementary to the different strands of each specific

sequence to be amplified. The primers must be sufficiently complem ntary to hybridiz with their respective strands. Therefore, the primer sequence need not reflect the exact sequence of the template. For example, a non-complementary nucleotide fragment may be attached to the 5' end of the primer, with the remainder of the primer sequence being complementary to the strand. Alternatively, noncomplementary bases or longer sequences can be interspersed into the primer, provided that the primer sequence has sufficient complementarity with the sequence of the strand to be amplified to hybridize therewith and thereby form a template for synthesis of the extension product of the other primer.

As used herein, a "bivalent primer" is a polynucleotide having two regions of complementarity to a predetermined target polynucleotide: (1) a first portion which is in the 5' portion of the bivalent primer and which is substantially complementary to a sequence in the 3' portion of the sequence to be amplified (target sequence) in the target polynucleotide, and (2) a second portion which is in the 3' portion of the primer and which is substantially complementary to a sequence in the 5' portion of the sequence to be amplified (target sequence) in the target polynucleotide. The portion of the bivalent primer which is substantially complementary to a sequence in the 3' portion of the sequence to be amplified (target sequence) is sufficiently long and sufficiently complementary to the target sequence to anneal under the reaction conditions and serve as an extendable primer for the polymerase to catalyze chain elongation. Similarly, the portion of the bivalent primer which is substantially complementary to a sequence in the 5' portion of the sequence to be amplified (target sequence) is sufficiently long and sufficiently complementary to the target sequence to anneal under the reaction conditions and serve as an extendable primer for the polymerase to catalyze chain elongation. Practitioners in the art will select at their discretion the specific structure of the bivalent primer(s) to be used in view of the necessity for annealing to the target. Typically, the portions of the bivalent primer which is

substantially complementary to a sequence in the 5' and 3' portions of the sequence to be amplified (target sequence) are each at least 12 to 15 nucleotid s in length, often 18 to 20 nucleotides in length, and are preferably 100 percent identical to the complement of the annealing portion of the target sequence. Often, bivalent primers of the invention are oligonucleotides.

The term "primer" as used herein refers to an oligonucleotide whether occurring naturally as in a purified restriction digest or produced synthetically, which is capable of acting as a point of initiation of synthesis when placed under conditions in which synthesis of a primer extension product which is complementary to a nucleic acid strand is induced, i.e., in the presence of nucleotides and an agent for polymerization such as DNA polymerase and at a suitable temperature and pH. The primer is preferably single-stranded for maximum efficiency in amplification, but may alternatively be double stranded. If double stranded, the primer is first treated to separate its strands before being used to prepare extension products. Preferably, the primer is an oligodeoxyribonucleotide. The primer must be sufficiently long to prime the synthesis of extension products in the presence of the agent for polymerization. The exact lengths of the primers will depend on many factors, including temperature and source of primers. For example, depending on the complexity of the target sequence, the oligonucleotide primer typically contains 15-25 or more nucleotides, although it may contain fewer nucleotides. Short primer molecules generally require cooler temperatures to form sufficiently stable hybrid complexes with template. In some embodiments, the primers can be large polynucleotides, such as from about 200 nucleotides to several kilobases or more.

As used herein, "suitable reaction conditions" are those conditions suitable for conducting PCR amplification using conventional reagents. Such conditions are known or readily established by thos of skill in the art, and can be exemplified by the r action conditions used in U.S. Patents 4,683,202, 4,683,195, and 4,800,159, which are incorporated

herein by reference. As one example and not to limit the invention, suitabl reaction conditions can comprise: 0.2mM each dNTP, 2.2 mM $MgCl_2$, 50 mM KCl, 10 mM Tris-HCl pH 9.0, 0.1% Triton X-100.

As used herein the term "physiological conditions" refers to temperature, pH, ionic strength, viscosity, and like biochemical parameters which are compatible with a viable organism, and/or which typically exist intracellularly in a viable cultured yeast cell or mammalian cell. For example, the intracellular conditions in a yeast cell grown under typical laboratory culture conditions are physiological conditions. Suitable in vitro reaction conditions for PCR and many polynucleotide enzymatic reactions and manipulations are generally physiological conditions. In general, in vitro physiological conditions comprise 50-200 mM NaCl or KCl, pH 6.5-8.5, 20-45°C and 0.001-10 mM divalent cation (e.g., Mg^{++}, Ca^{++}); preferably about 150 mM NaCl or KCl, pH 7.2-7.6, 5 mM divalent cation, and often include 0.01-1.0 percent nonspecific protein (e.g., BSA). A non-ionic detergent (Tween, NP-40, Triton X-100) can often be present, usually at about 0.001 to 2%, typically 0.05-0.2% (v/v). Particular aqueous conditions may be selected by the practitioner according to conventional methods. For general guidance, the following buffered aqueous conditions may be applicable: 10-250 mM NaCl, 5-50 mM Tris HCl, pH 5-8, with optional addition of divalent cation(s) and/or metal chelators and/or nonionic detergents and/or membrane fractions and/or antifoam agents and/or scintillants.

As used herein, the terms "label" or "labeled" refers to incorporation of a detectable marker, e.g., by incorporation of a radiolabeled nucleotide or incorporation of nucleotide having biotinyl moieties that can be detected by marked avidin (e.g., streptavidin containing a fluorescent marker or enzymatic activity that can be detected by optical or calorimetric methods). Various methods of labeling polynucleotides are known in the art and may be used. Examples of labels include, but are not limited to, the following: radioisotopes (e.g., ^{3}H, ^{14}C, ^{35}S, ^{125}I, ^{131}I),

fluorescent labels (e.g., FITC, rhodamine, lanthanide phosphors), enzymatic labels (e.g., horseradish peroxidase, β-galactosidase, luciferase, alkaline phosphatase), biotinyl groups, and the like. In some embodiments, labels are attached by spacer arms of various lengths to reduce potential steric hindrance.

As used herein, "substantially pure" means an object species is the predominant species present (i.e., on a molar basis it is more abundant than any other individual macromolecular species in the composition), and preferably a substantially purified fraction is a composition wherein the object species comprises at least about 50 percent (on a molar basis) of all macromolecular species present. Generally, a substantially pure composition will comprise more than about 80 to 90 percent of all macromolecular species present in the composition. Most preferably, the object species is purified to essential homogeneity (contaminant species cannot be detected in the composition by conventional detection methods) wherein the composition consists essentially of a single macromolecular species. Solvent species, small molecules (<500 Daltons), and elemental ion species are not considered macromolecular species.

The term "recombinant" used herein refers to macromolecules produced by recombinant DNA techniques wherein the gene coding for a polypeptide is cloned by known recombinant DNA technology. For example, an amplified or assembled product polynucleotide may be inserted into a suitable DNA vector, such as a bacterial plasmid, and the plasmid used to transform a suitable host. The gene is then expressed in the host to produce the recombinant protein. The transformed host may be prokaryotic or eukaryotic, including mammalian, yeast, Aspergillus and insect cells. One preferred embodiment employs bacterial cells as the host. Alternatively, the product polynucleotide may serve a non-coding function (e.g., promoter, origin of replication, ribosome-binding site, etc.).

Generally, the nomenclature used hereafter and many of the laboratory procedures in cell culture, molecular

genetics, and nucleic acid chemistry and hybridization described below are those well known and commonly employed in the art. Standard techniques are used for recombinant nucleic acid methods, polynucleotide synthesis, in vitro polypeptide synthesis, and the like and microbial culture and transformation (e.g., electroporation). Generally enzymatic reactions and purification steps are performed according to the manufacturer's specifications. The techniques and procedures are generally performed according to conventional methods in the art and various general references (see, generally, Sambrook et al. Molecular Cloning: A Laboratory Manual, 2d ed. (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.; each of which is incorporated herein by reference) which are provided throughout this document. The procedures therein are believed to be well known in the art and are provided for the convenience of the reader. All the information contained therein is incorporated herein by reference.

Oligonucleotides can be synthesized on an Applied Bio Systems oligonucleotide synthesizer according to specifications provided by the manufacturer.

Methods for PCR amplification are described in the art (PCR Technology: Principles and Applications for DNA Amplification ed. HA Erlich, Stockton Press, New York, NY (1989); PCR Protocols: A Guide to Methods and Applications, eds. Innis, Gelfland, Snisky, and White, Academic Press, San Diego, CA (1990); Mattila et al. (1991) Nucleic Acids Res. 19: 4967; Eckert, K.A. and Kunkel, T.A. (1991) PCR Methods and Applications 1: 17; and U.S. Patent Nos. 4,683,202 and 4,965,188, each of which are incorporated herein by reference) and exemplified hereinbelow.

Overview

A basis of the present invention is the use of polymerase in combination with at least two polynucleotides having complementary ends which can anneal whereby at least one of said polynucleotides has a free 3'-hydroxyl capable of polynucleotide chain elongation by a DNA polymerase, such as a

thermostable polymerase (e.g., Thermus aquaticus (Taq) polymerase or Thermococcus litoralis (Vent™) polymerase. In an embodiment, the method is performed using PCR, typically with multiple cycles of heat denaturation and DNA synthesis. However, there are several variations of the basic method of end-complementary polymerase reaction which are exemplified hereinbelow and which shall be evident to the skilled artisan in view of the present specification. Some variations do not require primers and/or sequential cycles of thermal denaturation.

In embodiments where the product size increases with the number of denaturation, annealing, and extension cycles (e.g., as the mean length of concatemers increase), it is typically advantageous to increase the denaturation temperature, and optionally increase the reannealing time, for subsequent cycles. Such conditions are readily optimized by the practitioner using pilot reactions to establish a calibration curve for any particular embodiment.

Single-Primer Amplification

A target polynucleotide is contacted with a bivalent primer typically comprising an oligonucleotide having two regions of complementarity to the target polynucleotide: (1) a first portion which is in the 5' portion of the primer and which is substantially complementary to a sequence in the 3' portion of the sequence to be amplified (target sequence) in the target polynucleotide, and (2) a second portion which is in the 3' portion of the primer and which is substantially complementary to a sequence in the 5' portion of the sequence to be amplified (target sequence) in the target polynucleotide. The contacting is performed under conditions suitable for hybridization of the bivalent primer to the target polynucleotide for polymerase-mediated chain elongation, most often following thermal denaturation of the target polynucleotide if it is initially present as a double-stranded form.

The first portion of the bivalent primer which is in the 5' portion of the primer and which is substantially

complementary to a sequence in the 3' portion of the sequence to be amplified (target sequence) in the target polynucleotide is typically at least 12 nucleotides in length, often at least 15 nucleotides in length, frequently at least 18 nucleotides in length, and is commonly 20 to 25 or more nucleotides in length, but usually does not exceed 10,000 nucleotides in length and is frequently less than 50 to 500 nucleotides in length. The first portion of the bivalent primer is substantially identical to the complement of a sequence at the 3' end of the target sequence, however there may be additional terminal nucleotides of the first portion of the bivalent primer which are substantially non-identical to a target sequence or its complement. Such terminal nucleotides must be substantially non-interfering so that their presence does not significantly inhibit the capability of the bivalent primer to selectively anneal to the target sequence and initiate chain elongation under suitable reaction conditions in the presence of polymerase. Although the first portion of the bivalent primer is substantially identical to the complement of a sequence at the 3' end of the target sequence, it need not be exactly identical; often a sequence identity of at least 80 percent is sufficient, typically at least 90 percent sequence identity is present, and preferably at least 95 percent or 100 percent sequence identity is present. As the length of the complementary sequence increases, typically the percentage of sequence identity necessary for specific annealing decreases within certain limits (pp. 399-407, in Berger and Kimmel, Methods in Enzymology, Volume 152, Guide to Molecular Cloning Techniques (1987), Academic Press, Inc., San Diego, CA, which is incorporated herein by reference).

The second portion of the bivalent primer which is in the 3' portion of the primer and which is substantially complementary to a sequence in the 5' portion of the sequence to be amplified (target sequence) in the target polynucleotide is typically at least 12 nucleotides in length, often at least 15 nucleotides in length, frequently at least 18 nucleotides in length, and is commonly 20 to 25 or more nucleotides in length, but usually does not exceed 10,000 nucleotides in

l ngth and is frequently less than 50 to 500 nucleotides in length. The first portion of the bivalent primer is substantially identical to the complement of a s quence at the 5' end of the target sequence, however there may be additional terminal nucleotides of the first portion of the bivalent primer which are substantially non-identical to a target sequence or its complement. Such terminal nucleotides must be substantially non-interfering so that their presence does not significantly inhibit the capability of the bivalent primer to selectively anneal to the target sequence and initiate chain elongation under suitable reaction conditions in the presence of polymerase. Although the first portion of the bivalent primer is substantially identical to the complement of a sequence at the 5' end of the target sequence, it need not be exactly identical; often a sequence identity of at least 80 percent is sufficient, typically at least 90 percent sequence identity is present, and preferably at least 95 percent or 100 percent sequence identity is present. In some embodiments, sequence identity of less than 80 percent is practicable, but the amount of sequence identity and length of overlap for the joints is determined by the discretion of the practitioner.

The amount of sequence identity necessary for any given application will vary depending on several factors including: (1) complexity of the population of polynucleotides in which the target polynucleotide(s) is/are present, (2) temperature and ionic strength, (3) sequence composition of the target sequence, (4) length of sequence identity, and (5) size of the primer. Practitioners will select bivalent primers having a first portion with sufficient sequence identity and length to serve as selective amplification primers which specifically hybridize to the desired target polynucleotide(s). Specific hybridization is the formation of hybrids between a primer polynucleotide and a target polynucleotide, wherein the primer polynucleotide preferentially hybridizes to the target DNA such that, for example, at least one discrete band can be identified on a gel of amplification products obtained from amplification of genomic DNA prepared from eukaryotic cells that contain (or

are spiked with) the target polynucl otide s quence. In some instances, a target sequence may be present in more than one target polynucleotide species (e.g., a particular target sequence may occur in multiple members of a gene family or in a known repetitive sequence). It is evident that optimal hybridization conditions will vary depending upon the sequence composition and length(s) of the targeting polynucleotide(s) and target(s), and the experimental method selected by the practitioner. Various guidelines may be used to select appropriate primer sequences and hybridization conditions (see, Maniatis et al., Molecular Cloning: A Laboratory Manual (1989), 2nd Ed., Cold Spring Harbor, N.Y.; Berger and Kimmel, Methods in Enzymology, Volume 152, Guide to Molecular Cloning Techniques (1987), Academic Press, Inc., San Diego, CA; PCR Protocols: A Guide to Methods and Applications, eds. Innis, Gelfland, Snisky, and White, Academic Press, San Diego, CA (1990); Benton WD and Davis RW (1977) Science 196: 180; Goodspeed et al. (1989) Gene 76: 1; Dunn et al. (1989) J. Biol. Chem. 264: 13057 which are incorporated herein by reference.

The target polynucleotide may be substantially homogeneous or may be present in a mixture of polynucleotide species (e.g., in a genome, biological sample, or mixture of synthetic polynucleotides). Subsequent or concomitant with the contacting of the bivalent primer to the target polynucleotide, a polynucleotide polymerase, such as a thermostable DNA polymerase, e.g., Taq polymerase, TThI polymerase (Perkin Elmer) or Vent™ (New England Biolabs, Beverly, MA), catalyzes, under suitable reaction conditions, polynucleotide synthesis (chain elongation) primed from the 3'-hydroxyl of the annealed bivalent primer to form a strand complementary to the target sequence, thereby forming a nascent complementary strand. Following completion of the nascent complementary strand spanning the target sequence, the target polynucleotide and the nascent strand are denatured, typically by elevation of temperature, and allowed to reanneal, typically by reduction of temperature, with another molecule of the bivalent primer species or with a

complementary strand of a target polynucleotide or an amplifi d copy thereof. The denatured nascent strand species following the first elongation cycle will contain a copy of the target sequence and has a terminal repeat of its 5'-terminal sequence at the 3' terminus, resulting from the bivalent primer, and wherein the terminal repeat is of sufficient length to support annealing under PCR conditions to an overlapping complementary strand in a head-to-tail arrangement (see, Fig. 1). Following reannealing, the described polymerase elongation/denaturation/reannealing cycle is repeated from 1 to about 100 times as desired, resulting in formation of amplified product which comprises head-to-tail concatemers of the target sequence. The concatemers typically increase in length as the number of amplification cycles increase and as the amount of bivalent primer decreases. Following amplification forming concatameric head-to-tail repeats of the target sequence, the concatemer(s) can optionally be resolved, such as (1) by cleaving with a restriction endonuclease which cuts within (or at the termini of) the concatemeric unit(s), (2) by homologous recombination between concatemer units to form covalently closed circles, or (3) by cleavage with a restriction endonuclease followed by ligation with DNA ligase to form covalently closed circles and/or by direct transformation into host cells for in vivo ligation.

By this method, a single primer (bivalent primer) is used to amplify a target polynucleotide sequence having a predetermined 5' terminal sequence and a predetermined 3' terminal sequence. The predetermined 5' terminal sequence and a predetermined 3' terminal sequence may be contained internally within a larger polynucleotide; hence the use of the term "terminal" refers only to their terminality within the target sequence, not necessarily the complete target polynucleotide which may be a superset of the target sequence.

Rolling Circle PCR Amplification

Often, a target polynucleotide sequence which is amplified by the present method will form amplification

intermediates in the form of cyclized DNA (see, Fig. 2), as a result of the 3' terminus of an overlapped nascent strand annealing to the 3' terminus of an overlapped complementary strand forming a cyclized (circular) structure similar to a gapped circle. The cyclized structure has a strand with an extendable 3'-hydroxyl which can be extended with a DNA polymerase substantially lacking exonuclease activity (e.g., a thermostable polymerase such as Vent(exo$^-$)^{TM} or Klenow fragment, etc.) in a rolling circle format whereby the leading terminus of the nascent strand continually displaces the lagging portion of the nascent strand (see, Fig. 2) producing a concatemeric single strand propagating from the rolling circle intermediate. Most often, such rolling circle intermediates will form under dilute conditions more favorable to intramolecular cyclization of overlapped strands than to formation of additional intermolecular overlaps. Once a rolling circle intermediate is established, the template need not be denatured in order to continue amplification of the target sequence as in conventional PCR, since the polymerase continues around the circle processively. Thus, the advantageous formation of the rolling circle intermediate in the present method avoids the necessity of multiple thermal cycles of PCR to repeatedly denature and renature the amplification template (and the resultant time loss needed for heating and cooling).

Overlapped Assembly of Polynucleotides

The present invention also provides for assembly of one or more product polynucleotide(s) from a plurality of component polynucleotides which have overlapping complementary sequence portions at their termini. The component polynucleotides are conveniently single-stranded oligonucleotides, but can include double-stranded polynucleotides (which are generally denatured with elevated temperature) and long single-stranded polynucleotides.

A desired product polynucleotide (or polynucleotide library) is assembled from a plurality of component polynucleotides by formation of overlapped strands of

alternating polarity and having substantially complementary termini (see, Fig. 3). This method employs a series of overlapping substantially complementary termini to determine the linear order of component sequences in the final product. Concomitant with or subsequent to formation of the overlapped strands of the component polynucleotides in a reaction, a polynucleotide polymerase (e.g., a thermostable DNA polymerase) under suitable reaction conditions catalyzes strand elongation from the 3'-hydroxyl portions of the overlapped (annealed) joints, filling in the portion between joints and processively displacing or processively degrading exonucleolytically the 5' termini of downstream component strands of the same polarity as the nascent strand elongates. After a cycle of chain elongation forming substantially double-stranded polynucleotides, the reaction conditions are altered (typically by increasing the temperature) to effect denaturation of the double-stranded polynucleotides, followed by altering the reaction conditions to permit reannealing of complementary strands or portions thereof (i.e., overlapping termini) to form molecules having overlapped termini (joints), and chain elongation by a polynucleotide polymerase under suitable reaction conditions catalyzes strand elongation from the 3'-hydroxyl portions of the overlapped (annealed) joints, as in the first cycle. One to about 100 cycles of denaturation/annealing/polymerization can be performed to generate a product comprising the component polynucleotide sequences covalently linked in linear order according to the order of the overlapping joints. In this embodiment, a product polynucleotide can be constructed from a plurality of smaller component polynucleotides (typically oligonucleotides) and enables assembly of a variety of products with alternate substitutable polynucleotide components at a given position serving as structural "alleles" (see, Fig. 4). The component polynucleotides are often provided in single-strand form, but may initially be present in double-strand form and be denatured (typically by elevated temperature) for the assembly of the product by PCR amplification. Substantially any type of product polynucleotide can be assembled in this way,

including cloning and expression vectors, viral genomes, gene therapy vectors, g n s (including chimeric genes), polynucleotides encoding peptide libraries, and the like. In a variation, one or more of the component polynucleotides represents a site-directed mutation or variable-sequence kernal. In a variation, PCR employing a low-fidelity polymerase is used to introduce additional sequence variation into the product polynucleotide(s) during amplification cycles. The method can be used to produce a library of sequence-variant product polynucleotides, if desired.

Kits

The invention also provides kits comprising a bivalent primer polynucleotide and/or a plurality of component polynucleotides and instructions for use describing the present end-complementary amplification method disclosed herein. Frequently, a polynucleotide polymerase, such as a thermostable DNA polymerase (Taq or Vent™ polymerase) is also present in the kit. Optionally, one or more target polynucleotides may be provided in the kit, such as for calibration and/or for use as a positive control to verify correct performance of the kit.

General Aspects

The target polynucleotides or component polynucleotides may be obtained from any source, for example, from plasmids such as pBR322, from cloned DNA or RNA, or from natural DNA or RNA from any source, including bacteria, yeast, viruses, and higher organisms such as plants or animals. DNA or RNA may be extracted from blood, tissue material such as chorionic villi or amniotic cells by a variety of techniques such as that described by Maniatis et al., Molecular Cloning: A Laboratory Manual, (New York: Cold Spring Harbor Laboratory, 1982), pp. 280-281. Alternatively, the polynucleotides may be produced by chemical synthesis by any of the art-recognized methods.

Any specific nucleic acid sequence can be produced by the present process. It is only necessary that a sufficient

number of bases at both ends of the sequence be known in sufficient detail so that a bivalent primers can be prepared which will hybridize to the desired sequence and at relative positions along the sequence such that an extension product initially synthesized from the bivalent primer, when it is separated from its template (complement), can anneal with a stand of the opposite polarity to form an overlapped joint of a head-to-tail concatemer and serve as a template for extension of the 3'-hydroxyl from each overlapped joint. The greater the knowledge about the bases at both ends of the sequence, the greater can be the specificity of the primer for the target nucleic acid sequence, and thus the greater the efficiency of the process. It will be understood that the word bivalent primer as used hereinafter may refer to more than one bivalent primer, particularly in the case where there is some ambiguity in the information regarding the terminal sequence(s) of the fragment to be amplified. For instance, in the case where a nucleic acid sequence is inferred from protein sequence information a collection of primers containing sequences representing all possible codon variations based on degeneracy of the genetic code will be used for each strand.

The polynucleotide primers may be prepared using any suitable method, such as, for example, the phosphotriester and phosphodiester methods, or automated embodiments thereof. In one such automated embodiment diethylphosphoramidites are used as starting materials and may be synthesized as described by Beaucage et al. (1981) Tetrahedron Letters 22: 1859. One method for synthesizing oligonucleotides on a modified solid support is described in U.S. Pat. No. 4,458,066. It is also possible to use a primer which has been isolated from a biological source (such as a restriction endonuclease digest or the like.

The specific nucleic acid sequence is produced by using the target polynucleotide containing that sequence as a template. If the target polynucleotide contains two strands, it is necessary to separate the strands of the nucleic acid before it can be used as the template, either as a separate

step or simultaneously with the synthesis of the primer extension products. This strand separation can be accomplished by any suitable denaturing method including physical, chemical or enzymatic means. One physical method of separating the strands of the polynucleotide involves heating the polynucleotide until it is substantially denatured. Typical heat denaturation may involve temperatures ranging from about 80° to 105° C for times ranging from about 10 seconds to about 10 minutes or more. Strand separation may also be induced by an enzyme from the class of enzymes known as helicases or the enzyme RecA, which has helicase activity and in the presence of rATP is known to denature DNA. The reaction conditions suitable for separating the strands of polynucleotides with helicases are described by Cold Spring Harbor Symposia on Quantitative Biology, Vol. XLIII "DNA: Replication and Recombination" (New York: Cold Spring Harbor Laboratory, 1978), B. Kuhn et al., "DNA Helicases", pp. 63-67, and techniques for using RecA are reviewed in C. Radding, Ann. Rev. Genetics, 16:405-37 (1982).

PCR synthesis can be performed using any suitable method. Generally it occurs in a buffered aqueous solution, preferably at a pH of 7-9, most preferably about 8. The bivalent primer(s) is/are added in suitable amounts (molar ratio to target), typically less than conventional PCR methods because of the self-priming nature of the overlapped concatemers. The deoxyribonucleoside triphosphates dATP, dCTP, dGTP and TTP are also added to the synthesis mixture in adequate amounts and the resulting solution is heated to about 85°-100° C. for from about 1 to 10 minutes, preferably from 1 to 4 minutes. After this heating period the solution is allowed to cool to from 20°-40° C, which is preferable for the primer hybridization. To the cooled mixture is added an agent for polymerization, and the reaction is allowed to occur under conditions known in the art. This synthesis reaction may occur at from room temperature up to a temperature above which the agent for polymerization no longer functions efficiently. Thus, for example, if DNA polymerase is used as the agent for polymerization, the temperature is generally no greater than

about 45° C. The agent for polymerization may be any compound or system which will function to accomplish the synthesis of primer extension products, including enzymes. Suitabl enzymes for this purpose include, for example, E. coli DNA polymerase I, Klenow fragment of E. coli DNA polymerase I, T4 DNA polymerase, other available DNA polymerases, reverse transcriptase, and other enzymes, including heat-stable enzymes, which will facilitate combination of the nucleotides in the proper manner to form the primer extension products which are complementary to each nucleic acid strand. Generally, the synthesis will be initiated at the 3' end of the primer and proceed in the 5' direction along the template strand, until synthesis terminates, producing molecules of different lengths.

The newly synthesized strand and its complementary nucleic acid strand form a double-stranded molecule which is used in the succeeding steps of the process. In the next step, the strands of the double-stranded molecule are separated using any of the procedures described above to provide single-stranded molecules.

The steps of strand separation and extension product synthesis can be repeated as often as needed to produce the desired quantity of the specific nucleic acid sequence. The amount of the specific nucleic acid sequence produced will accumulate in an exponential fashion and the average size of the product will also increase as the length of the concatemers increases with each cycle.

The method herein may also be used to enable detection and/or characterization of specific nucleic acid sequences associated with infectious diseases, genetic disorders or cellular disorders such as cancer, e.g., oncogenes. Amplification is useful when the amount of nucleic acid available for analysis is very small, as, for example, in the prenatal diagnosis of sickle cell anemia using DNA obtained from fetal cells.

Continuous Multiplex Amplification

Continuous multiplex amplification can be used to amplify by any suitable amplification method, typically by PCR, a plurality of unlinked or distantly linked polynucleotide sequences. Certain genetic diagnostic tests require amplification of multiple segments (e.g., exons) of a gene. Each segment is typically amplified in a separate amplification reaction. Unfortunately, it is generally difficult or impossible to amplify each segment in approximately equimolar ratios due to differences in priming efficiency, length of extension, secondary structure, or other factors which affect amplification rate. In continuous multiplex amplification, the amplification reactions can be run together in a single reaction vessel using a common pool of reagents where the unlinked (or distantly linked) sequences become part of the same amplification product, which affords the substantially equimolar amplification of the unlinked (or distantly linked) sequences. An embodiment of the invention is illustrated schematically in Figs. 9A-9C in the case of amplifying two unlinked sequences represented in the double-stranded polynucleotide fragments ABC/A'B'C' and DEF/D'E'F'. primers C'X'D', F'X'A', FYA, and CYD are added and annealed to the denatured polynucleotide fragments; the primer concentrations are typically lower than conventionally used for PCR primers. X and Y, and their complements X' and Y' are generally predetermined sequences which are selected to destabilize the primer:primer hybrids CYD/C'X'D' and FYA/F'X'A', such as by having the X and Y sequences (and their complements) lack substantial sequence identity. After extension with a polymerase, the following products and their complements result: ABCXD, DEFXA, FYABC, and CYDEF. A variety of hybrid combinations of product:product and primer:product can form and after another round of amplification a variety of amplification products result. Each of the pairs is capable of self-priming with its complement or with the complement of another fragment which has a complementary sequence. Through multiple cycles of amplification, the initial primer population becomes depleted and primarily extended products remain. These extended products will prime each other and

generate increasingly longer amplification products which contain the initial unlinked (or distantly linked) s qu nces in equal amounts. After completion of amplification, several options can be pursued; the amplification product(s) can be used directly, the X and/or Y sequences can contain restriction sites (preferably unique sites) to allow digestion with the restrictino enzyme and, if desired, separation and/or purification of the two (or more) originally unlinked sequences. Alternatively, or in combination, transcription promoters (e.g., T3 and T7) can be included in X and/or Y sequences to facilitate transcription of the amplified sequences. Figs. 10A-10C show a linear format of the continuous multiplex amplification method. Figs. 11A-11C show an embodiment of circular continuous multiplex amplification wherein bivalent primers contain T3 and T7 promoters and the functional promoter sequences are thereby introduced into the amplification product(s).

The following examples are given to illustrate the invention, but are not to be limiting thereof.

EXPERIMENTAL EXAMPLES

The following examples are offered by way of example and not by way of limitation. Variations and alternate embodiments will be apparent to those of skill in the art.

Example 1. LacZ alpha gene reassembly

This example shows that small fragments having overlapping regions of homology can be amplified and reassembled by PCR amplification methods in the absence of any primer.

1) Substrate preparation

The substrate for the reassembly reaction was the dsDNA polymerase chain reaction ("PCR") product of the wild-type LacZ alpha gene from pUC18. (Gene Bank No. X02514) The primer sequences were 5'AAAGCGTCGATTTTTGTGAT3' (SEQ ID NO:1) and 5'ATGGGGTTCCGCGCACATTT3' (SEQ ID NO:2). The free primers were removed from the PCR product by Wizard PCR prep (Promega,

Madison WI) according to the manufacturer's directions. The removal of the free primers was found to be important.

2) DNAseI digestion

About 5 μg of the DNA substrate was digested with 0.15 units of DNAseI (Sigma, St. Louis MO) in 100 μl of (50 mM Tris-HCl pH 7.4, 1 mM $MgCl_2$), for 10-20 minutes at room temperature. The digested DNA was run on a 2% low melting point agarose gel. Fragments of 10-70 basepairs (bp) were purified from the 2% low melting point agarose gels by electrophoresis onto DE81 ion exchange paper (Whatman, Hillsborough OR). The DNA fragments were eluted from the paper with 1 M NaCl and ethanol precipitated.

3) DNA Reassembly

The purified fragments were resuspended at a concentration of 10 - 30 ng/μl in PCR Mix (0.2 mM each dNTP, 2.2 mM $MgCl_2$, 50 mM KCl, 10 mM Tris-HCl pH 9.0, 0.1% Triton X-100, 0.3 μl Taq DNA polymerase, 50 μl total volume). No primers were added at this point. A reassembly program of 94°C for 60 seconds, 30-45 cycles of [94°C for 30 seconds, 50-55°C for 30 seconds, 72°C for 30 seconds] and 5 minutes at 72°C was used in an MJ Research (Watertown MA) PTC-150 thermocycler. The PCR reassembly of small fragments into larger sequences was followed by taking samples of the reaction after 25, 30, 35, 40, and 45 cycles of reassembly.

Whereas the reassembly of 100-200 bp fragments can yield a single PCR product of the correct size, 10-50 base fragments typically yield some product of the correct size, as well as products of heterogeneous molecular weights. Most of this size heterogeneity appears to be due to single-stranded sequences at the ends of the products, since after restriction enzyme digestion a single band of the correct size is obtained.

4) PCR with primers

After dilution of the r assembly product into the PCR Mix with 0.8 μM of each of the above primers (SEQ ID Nos: 1 and 2) and about 15 cycl s of PCR, each cycle consisting of (94°C for

30 seconds, 50°C for 30 seconds and 72°C for 30 seconds), a single product of the correct size was obtained.

5) Cloning and analysis

The PCR product from step 4 above was digested with the terminal restriction enzymes *Bam*HI and *Eco*0109 and gel purified as described above in step 2. The reassembled fragments were ligated into pUC18 digested with *Bam*HI and *Eco*0109. *E. coli* were transformed with the ligation mixture under standard conditions as recommended by the manufacturer (Stratagene, San Diego CA) and plated on agar plates having 100 μg/ml ampicillin, 0.004% X-gal and 2mM IPTG. The resulting colonies having the *Hin*DIII-*Nhe*I fragment which is diagnostic for the ++ recombinant were identified because they appeared blue.

This Example illustrates that a 1.0 kb sequence carrying the LacZ alpha gene can be digested into 10-70 bp fragments, and that these gel purified 10-70 bp fragments can be reassembled to a single product of the correct size, such that 84% (N=377) of the resulting colonies are $LacZ^+$ (versus 94% without shuffling). This principal finding is extended substantially in the present invention to assemble component polynucleotides into product polynucleotides, and the component polynucleotides are not limited to randomly digested fragments of a naturally-occurring gene sequence.

The DNA encoding the LacZ gene from the resulting $LacZ^-$ colonies was sequenced with a sequencing kit (United States Biochemical Co., Cleveland OH) according to the manufacturer's instructions and the genes were found to have point mutations due to the reassembly process (Table 1). 11/12 types of substitutions were found, and no frameshifts.

TABLE 1

Mutations introduced by mutagenic shuffling

Transitions	Frequency	Transversions	Frequency
G - A	6	A - T	1
A - G	4	A - C	2
C - T	7	C - A	1
T - C	3	C - G	0
		G - C	3
		G - T	2
		T - A	1
		T - G	2

A total of 4,437 bases of shuffled lacZ DNA were sequenced.

The rate of point mutagenesis during DNA reassembly from 10-70 bp pieces was determined from DNA sequencing to be 0.7 % (N=4,473), which is similar to error-prone PCR. Without being limited to any theory it is believed that the rate of point mutagenesis may be lower if larger fragments are used for the reassembly, or if a proofreading polymerase is added.

When plasmid DNA from 14 of these point-mutated $LacZ^-$ colonies were combined and again reassembled/shuffled by the method described above, 34% (N=291) of the resulting colonies were $LacZ^+$, and these colonies presumably arose by recombination of the DNA from different colonies.

The expected rate of reversal of a single point mutation by error-prone PCR, assuming a mutagenesis rate of 0.7% (10), would be expected to be <1%.

Thus large DNA sequences can be reassembled from a random mixture of small fragments by a reaction that is surprisingly efficient and simple. One application of this technique is the recombination or shuffling of related sequences based on homology. A second application is the assembly of a large product polynucleotide by PCR amplification of component polynucleotides (oligonucleotides) having overlapping regions of homology to form annealed joints during PCR amplification.

Example 2. One-Step Circular Plasmid Assembly From Oligonucleotides

This example demonstrates assembly of a 2.71 kb plasmid p182SfiI (Stemmer (1994) Nature 370: 389) which encodes the gene and promoter region for R-TEM1 β-lactamase.

A collection of 132 component oligonucleotides, each 40 bases in length, as well as one 56-mer and one 47-mer (see, Fig. 5A-E) were synthesized and used to assemble the plasmid by end-complementary polymerase reaction (ECPR) employing the overlapping ends of the oligonucleotides. This collection of component polynucleotides collectively encode the plasmid p182SfiI. The plus strand and the minus strand were each initially directed by oligonucleotides 40 nucleotides long which, upon assembly, overlapped by 20 nucleotides (Fig. 6). The oligonucleotides were synthesized and 5'-phosphorylated simultaneously on a 96-well parallel-array DNA synthesizer using standard phosphoramidite chemistry. After cleavage from the solid support and deprotection, the dried down oligonucleotides were resuspended in distilled water and used without further purification.

The oligonucleotides were diluted to a final concentration (all oligos combined) of 1μM (14 ng/μl) in 20 μl of GeneAMP XL PCR Mix (Perkin-Elmer, Branchburg, NJ; 0.2mM each dNTP, 2.2 mM $MgCl_2$, 50 mM KCl, 10 mM Tris-HCl pH 9.0, 0.1% Triton X-100). An aliquot of the reaction mix (A) was electrophoresed on an agarose gel (Fig. 7). The reaction was started with 5 μl of a 50:1 (v/v) mixture of Taq polymerase (Promega, Madison, WI) and Pfu polymerase (Stratagene, La Jolla, CA) such that 1 unit of Taq and 0.02 unit of Pfu polymerase were added. The PCR program consisted of the following program: 40°C for 2 minutes, 72°C for 10 seconds, then 40 cycles of (94°C for 15 seconds, 40°C for 30 seconds, and 72°C for [10 seconds + 1 second/cycle]). An aliquot of the resulting reaction product (B) was electrophoresed on an agarose gel (Fig. 7), the remainder was then diluted 3x with XL PCR Mix and enzyme and amplified with the following program: 25 cycles of (94°C for 15 seconds, 40°C for 30 seconds, and 72°C for [45 seconds + 1 second/cycle]). An aliquot of the resulting reaction product (C) was electrophoresed on an agarose gel (Fig. 7) and the remainder then diluted 3x with XL PCR Mix and enzyme and amplified with the following program: 20 cycles of (94°C for 15 seconds, 40°C for 30 seconds, and 72°C for [70 seconds + 1 second/cycle]).

An aliquot of the resulting reaction product (D) was electrophoresed on an agarose gel (Fig. 7).

The an aliquot of the reaction product (D) was electrophoresed on an agarose gel, as was an aliquot which was digested with BamHI, which verified assembly of large DNA molecules consistent with formation of large concatemers which was resolved to unit length with BamHI digestion. The PCR product obtained by this method was shown to be concatemeric, and was resolvable by BamHI digestion into a single 2.7 kb band by agarose gel electrophoresis.

Aliquots of reaction product (D) were digested with various restriction enzymes (shown in Fig. 7). Fig. 7 shows that the unique cutters EcoRI and BamHI liberated a 2.71 kb fragment consistent with the size of a complete 2.71 kb plasmid p182SfiI (Stemmer (1994) Nature 370: 389, incorporated herein by reference). Furthermore, the digestion results with NcoI, SfiI, PstI, and BglII all yielded fragments consistent with the restriction map of the complete 2.71 kb plasmid p182SfiI.

After digestion of the PCR product with BamHI, the 2.7 kb fragment was gel purified and ligated with ligase, and transformed into E. coli.and transformed into E. coli K-12. Tetracycline-resistant transformants were selected.

This example demonstrates that the circular DNA assembly method allows for rapid and inexpensive construction of long DNA sequences, such as genes, gene libraries, plasmids, viral genomes, etc. The assembly method facilitates several mutagenesis approaches, such as point mutagenesis, combinatorial cassette mutagenesis, and doping, or mixing in other nucleotides during oligonucleotide synthesis. Deliberate modifications to the DNA sequence can be made simply by substituting one or more new oligos followed by reassembly. To reduce the rate of PCR mutagenesis during assembly, the addition of a proofreading polymerase can assure efficient long-read PCR reactions by combining high processivity with proofreading.

Example 3. Antibody Germline Assembly from Oligos With Rolling Circle Concatemeric Amplification

A scFv antibody with germline sequences (V_H251 and V_{IA}25) was constructed from 19 oligonucleotides by cyclized assembly. The oligos were at 2-20 ng per μl in PCR Mix. the program was 20 cycles of (94°C for 15s, 48°C for 30s, 72°C for 30+1s/cycle). The size of the product of this reaction was 200-500 bp. The PCR product was diluted 4-fold in PCR Mix and PCR was run for 24 cycles of (94°C for 15s, 55°C for 30s, 72°C for 30+8s/cycle), followed by one additional 3-fold dilution and 20 cycles of (94°C for 15s, 55°C for 30s, 72°C for 30+8s/cycle). The product was >50 kilobases, and after digestion with SfiI and NotI resulted in a single DNA fragment of the correct size.

Fig. 8 shows a schematic for end-complementary polymerase reaction (ECPR) in conjunction with parallel-processing PCR to amplify very large polynucleotides, such as those larger than can be amplified reliably by conventional PCR using only a single primer set.

Example 4. Plasmid Assembly With Rolling Circle Concatemeric Amplification

pGJ103 is a 5.5 kilobase plasmid containing an intact *ars* operon (Ji and Silver (1992) Proc. Natl. Acad. Sci. (USA) 89: 9474).

In one example, pGJ103 was digested with DNAseI into random 100-400 bp fragments which were reassembled by circular shuffling in PCR Mix with a program of 50 cycles of (94°C for 15s, 68°C for 30s+8s/cycle), using three different concentrations of fragments. Each reassembly yielded a product of >50 kb which was digested with BamHI to yield a single band of the correct (predicted) size, which was ligated, transformed into E. coli, and preplated on increasing levels of arsenate to select for up-mutants.

Cells and plasmids. Plasmid pGJ103 is pUC19 derivative containing the 2.5 kb arsenic resistance operon from S. aureus

plasmid pI258. E. coli strain TG1 was obtained commercially (Pharmacia, Tarrytown, NJ). Sodium arsenate (Sigma) was used as a 2.5M stock solution, neutralized to pH 7 with NaOH. Selection for increased arsenate resistance was performed at 35°C on agar plates with LB medium (Life Technologies) containing varying concentrations of arsenate.

The 5.5 kb plasmid pGJ103 was fragmented by sonication into fragments of 400-1500 bp, and reassembled by PCR using Perkin Elmer XL-PCR reagents with 10% PEG-6000, using PCR without added primers. The PCR program for the assembly was 90°C 30 s, than 60 cycles of: 94°C 20 s, 40-45°C 30 s, 72°C 39s + 1 s per cycle in a PTC-150 minicycler (MJ Research, Watertown, MA). The PCR process yielded plasmid multimers of about 15 to 40 kb in size, that were digested into 5.5 kb monomers with the restriction enzyme BamHI, which has a single unique site in plasmid pGJ103. The 5.5 kb plasmid monomer was purified from an agarose gel after electrophoresis, and after self-ligation, it was electroporated into electrocompetent E. coli TG1 cells.

Arsenate resistance selection. Transformed E. coli cells were plated on LB plates containing a range of concentrations of sodium and arsenate and incubated at 37°C for 24 hrs., and at least 1,000 colonies from the plates with the highest arsenate levels were pooled by scraping the plates. The harvested cells were grown in liquid in the presence of the same concentration of arsenate as in the petri dish and a plasmid pool was prepared from this liquid culture. Rounds 2-4 were identical to round 1, except that the cells were plated at higher arsenate levels.

Arsenate resistance quantification. Induced inoculum cells of E. coli TG1(pGJ103), wild type ars operon, and TG1 with mutant pGJ103 plasmid pools were grown overnight at 37°C in 2 mM or 50 mM arsenate, respectively. Equal amounts of cells (by turbidity as OD600nm) were on plates containing a range of concentrations of arsenate and grown for 18 hrs. at 37°C.

Cell growth was quantitated by r suspending the cells and measuring the OD600nm.

Arsenate detoxification assay. The ability of E. coli constructs to detoxify arsenate was measured by intact cell arsenate reduction assay using radioactive 73As043-as substrate and separation of arsenate and arsenite of by thin layer chromatography, followed by quantitation in an Ambis radioactive counter.

DNA sequencing. The sequence of the entire operon after selection was determined by dideoxy DNA sequencing using fluorescent terminating substrate and an ABI sequencer.

Results and discussion. The wild type plasmid pGJ103 ars operon confers on E. coli strain TG1 resistance to up to 4 mM arsenate when grown on LB plates at 37°C for 24 hrs. Selection round one, which was plated on 2, 4, 8, 16 and 32 mM arsenate, yielded about 2,000 colonies growing at 16 mM arsenate. Selection round two was plated 16, 32, 64 and 128 mM arsenate and yielded about 4,000 colonies growing at 64 mM arsenate. Round three was plated at 64, 128 and 256 mM arsenate and yielded about 1,500 colonies at 128 mM arsenate, and round 4 was plated on 128, 256 and 512 mM arsenate. Colonies were harvested from the plates with 256 mM arsenate and replated on 200, 300 and 400 mM arsenate. Single colonies from plates with 400 mM arsenate were grown in liquid culture with 400 mM arsenate and frozen at -70°C, and used for further characterization. Resistance levels were increased by DNA shuffling to arsenate (as selected) and also to arsenite and antimony salts (Fig. 12), which are the two toxic oxyanions to which resistance requires the ArsB membrane transporter but not the ArsC arsenate reductase enzyme. In this growth experiment, done with the pool from three cycles of DNA shuffling (which retained good growth in LB broth), not only was growth clear about 100 mM As043-, but increased resistance to arsenite (As02-) and antimony (Sb03+) was clearly=20 shown. These results require mutational effects beyond those possibly

limited to the arC gene, which affects resistance to arsenate alone.

Chromosomal integration. Cells selected and grown at and above 128 mM arsenate resulted in smaller growth, lower cell growth yields, and in low and variable plasmid yields. Plasmids were isolated that had apparently lost the arsenate operon, and most cells showed a complete loss of plasmids. It appeared that the DNA shuffling plus selection for high arsenate resistance resulted in integration of the ars operon into the E. coli chromosome, since the ars operon could be recovered from chromosomal DNA of clones which had lost the entire plasmid by conventional PCR amplification with "upstream" and "downstream" oligonucleotide primers.

Integration mechanism. The arsenate resistance operon of plasmid pGJ103 is flanked on both sides by 200 bp inverted homologous regions, which appear to be the terminal portions of site-specific recombinase genes. Attempts to recover the operon from the total cellular DNA of highly resistant cells by PCR showed that oligonucleotide primers near the inside ends of the recombinanse genes, immediately flanking the arsenate genes, yielded a PCR product of the correct size (2.0 kb) and with the expected restriction nuclease site pattern. However, primers located toward the middle or near the outside ends of the 200 bp homologous sequences did not yield predicted PCR products with the intact ars operon. Presumably, chromosomal integration was selected because the integrated operon somehow resulted in increased arsenate resistance and the homologous sequences at the ends of the ars operon facilitated chromosomal integration by recombination.

Chromosomal ars operon. The chromosome of E. coli normally contains an arsenate resistance operon which is distantly related to the pI258 operon and results in a low level of arsenate resistance. The operon which was recovered from the chromosome of highly resistant cells by PVR was shown by restriction mapping and by DNA sequencing to be derived from

the pI258 operon, and not from the E. coli K-12 chromosomal operon.

Chromosomal shuffling. Because the cells recovered from 128 mM arsenate did not contain plasmid DNA, the shuffling for round 4 was performed on the PCR product which was obtained from the chromosomal DNA of the cells selected in round 3. This PCR product was combined with a 10-fold lower molar amount of the plasmid DNA obtained from round 2 cells, and the mixture was fragmented, shuffled and selected as for earlier rounds.

Cloning and characterization of the integrated operon. The conventional PCR product which was obtained from the chromosomal DNA of cells grown at 400 mM arsenate was cloned into the polylinker site of pUC19. This construct was similar to pGJ103 except that it lacked the 200 bp inverted homologous DNA flanking the arsenate operon. Cells containing this plasmid were resistant only up to about 10 mM arsenate. The reason for this loss of arsenate resistance level is not known. The DNA sequence of this cloned chromosomal operon showed thirteen base changes relative to the original sequence. The arsR gene contained two silent mutations (T389C and T429C. The arsB gene contained ten base changes, and one base change occurred in the non-coding area past the end of the arsC gene (G2469C). Of the ten base changes in arsB, three resulted in amino acid alterations: base T1281C change resulted in amino acid change L232S, base T1317C change resulted in amino acid change F244S, and base T1853C change resulted in amino acid change Y423H, all three involving a change toward a more hydrophilic residue via a T to C transition. The seven silent mutations were T961G, A976G, T1267C, A1402G, T1730C, T1819C and T1844C.

Arsenate reductase activity. The activity of arsenate reductase by whole mutant cells after the third cycle was increased about 50-fold (Fig. 13) to the wild type initial strain with plasmid pGJ103. This increase in whole cell

reductase rate appeared to be more dependent on an increase in rate reduction and not on an enhanced affinity of arsenate for the cells (data not shown). This is consistent with the finding that the mutations occurred in the efflux transport protein and not in the arsenate reductase itself.

Although the present invention has been described in some detail by way of illustration for purposes of clarity of understanding, it will be apparent that certain changes and modifications may be practiced within the scope of the claims.

CLAIMS:

1. A method for amplifying a target polynucleotide, comprising:

contacting under conditions suitable for PCR, a target polynucleotide with a bivalent primer which comprises two portions of complementarity to the target polynucleotide: (1) a first portion which is in the 5' portion of the primer and which is substantially complementary to a sequence in the 3' portion of the sequence to be amplified (target sequence) in the target polynucleotide, and (2) a second portion which is in the 3' portion of the primer and which is substantially complementary to a sequence in the 5' portion of the sequence to be amplified (target sequence) in the target polynucleotide;

catalyzing under suitable reaction conditions for PCR, polynucleotide synthesis primed from the 3'-hydroxyl of the annealed bivalent primer to form a strand complementary to the target sequence, thereby forming a nascent complementary strand;

denaturing the target polynucleotide and the nascent strand and allowing reannealing with another molecule of the bivalent primer species or with a complementary strand of a target polynucleotide or an amplified copy thereof; and

repeating an elongation/denaturation/reannealing cycle from 1 to about 100 times as desired, resulting in formation of amplified product which comprises head-to-tail concatemers of the target sequence.

2. The method of claim 1, comprising the further step of cleaving said concatemers with a restriction endonuclease which cuts within each concatemeric unit to form a population of polynucleotides each consisting of an amplified target sequence.

3. The method of claim 2, comprising the further step of ligating the population of polynucleotides each

consisting of an amplifi d target sequence with DNA ligase to form covalently closed circles.

4. The method of claim 2, comprising the further step of ligating the population of polynucleotides each consisting of an amplified target sequence by direct transformation into host cells for in vivo ligation.

5. A method of claim 1, comprising the further step of annealing, under dilute conditions suitable for substantial intramolecular annealing and circle formation, the nascent strand with a complementary strand of a target polynucleotide or an amplified copy thereof to form amplification intermediates in the form of cyclized DNA as a result of the 3' terminus of an overlapped nascent strand annealing to the 3' terminus of an overlapped complementary strand which has a strand with an extendable 3'-hydroxyl which can be extended with a DNA polymerase substantially lacking exonuclease activity whereby the leading terminus of the nascent strand continually displaces the lagging portion of the nascent strand producing a concatemeric single strand emanating from the amplification intermediate.

6. A method for assembling a polynucleotide from a plurality of component polynucleotides, comprising:

a first step comprising contacting a plurality of strands of alternating polarity which comprise substantially complementary termini to form overlapped annealed joints in a reaction with a polynucleotide polymerase under suitable reaction conditions thereby catalyzing strand elongation from the 3'-hydroxyl portions of the overlapped joints, filling in a portion between said overlapped joints and processively displacing or processively degrading exonucleolytically the 5' termini of downstream component strands of the same polarity as the nascent strand elongates forming double-stranded polynucleotides;

a second step comprising denaturing of th double-stranded polynucleotides, followed by altering the reaction

conditions permitting reannealing of complementary strands or portions thereof to form concatemers having overlapped joints, and permitting chain elongation by a polynucleotide polymerase from the 3'-hydroxyl portions of the overlapped joints; and

repeating said second step from 1 to 100 times to generate a double-stranded concatemer of a desired size.

7. The method of claim 6, wherein the component polynucleotides are in single-strand form.

8. The method of claim 6, wherein the polymerase is Taq polymerase, TthI polymerase or Vent polymerase.

9. The method of claim 6, wherein the step of denaturation comprises elevating the temperature of the reaction to at least 94°C.

10. A kit comprising a bivalent primer and instructions for performing the method of claim 1.

11. A method for producing a polynucleotide by overlap assembly of parallel polymerase chain reaction amplifications, comprising:

forming at least three overlapping polynucleotides, wherein the 3' terminus of a first single-stranded polynucleotide is substantially complementary to the 3' terminus of a third single-stranded polynucleotide of the opposite polarity, and wherein the 5' terminus of said second single-stranded polynucleotide is substantially complementary to the 3' terminus of a third single-stranded polynucleotide having polarity identical to said first single-stranded polynucleotide, thereby generating an overlapped polynucleotide set capable of chain elongation by a suitable polymerase to generate a double-stranded product spanning the three initial overlapped polynucleotides;

synthesizing complementary strand polynucleotide sequences catalyzed by a DNA polymerase to form a double-stranded product spanning the overlapped polynucleotide set.

12. Th method of claim 11, wherein a plurality of primer sets are employed in a single reaction, each primer set priming the PCR amplification of a polynucleotid sequence which comprises terminal sequences which are complementary to terminal sequences in at least one other amplification product produced by a different primer set, thus generating a set of overlapping PCR products with which an overlapped polynucleotide set spanning the entire set of PCR products is generated and complementary sequences synthesized by polymerase to form a double-stranded polynucleotide spanning the entire set of overlapping PCR products.

13. The method of claim 12, comprising the further step of circularizing the double-stranded polynucleotide spanning the entire set of overlapping PCR products and ligating with ligase in vitro or transforming directly into suitable host cells.

14. The method of claim 12, wherein the double-stranded polynucleotide spanning the entire set of overlapping PCR products is at least 100 kilobases.

15. A method of continuous multiplex amplification wherein a plurality of initially unlinked polynucleotide sequences are amplified with at least two bivalent primers forming an amplifcaiton product comprising an equimolar ratio of the initially unlinked polynucleotide sequences and complementary ends.

16. The method of claim 15, wherein the multiplex amplification method is circular multiplex PCR amplification.

17. The method of claim 15, wherein the multiplex amplification method is linear multiplex PCR amplification.

18. The method of claim 15, wherein the multiplex amplification method results in amplification prosucts

comprising at least one prokaryotic or bacteriophage promoter sequence.

FIGURE 1A

A method to amplify single stranded DNA fragments using a single primer.

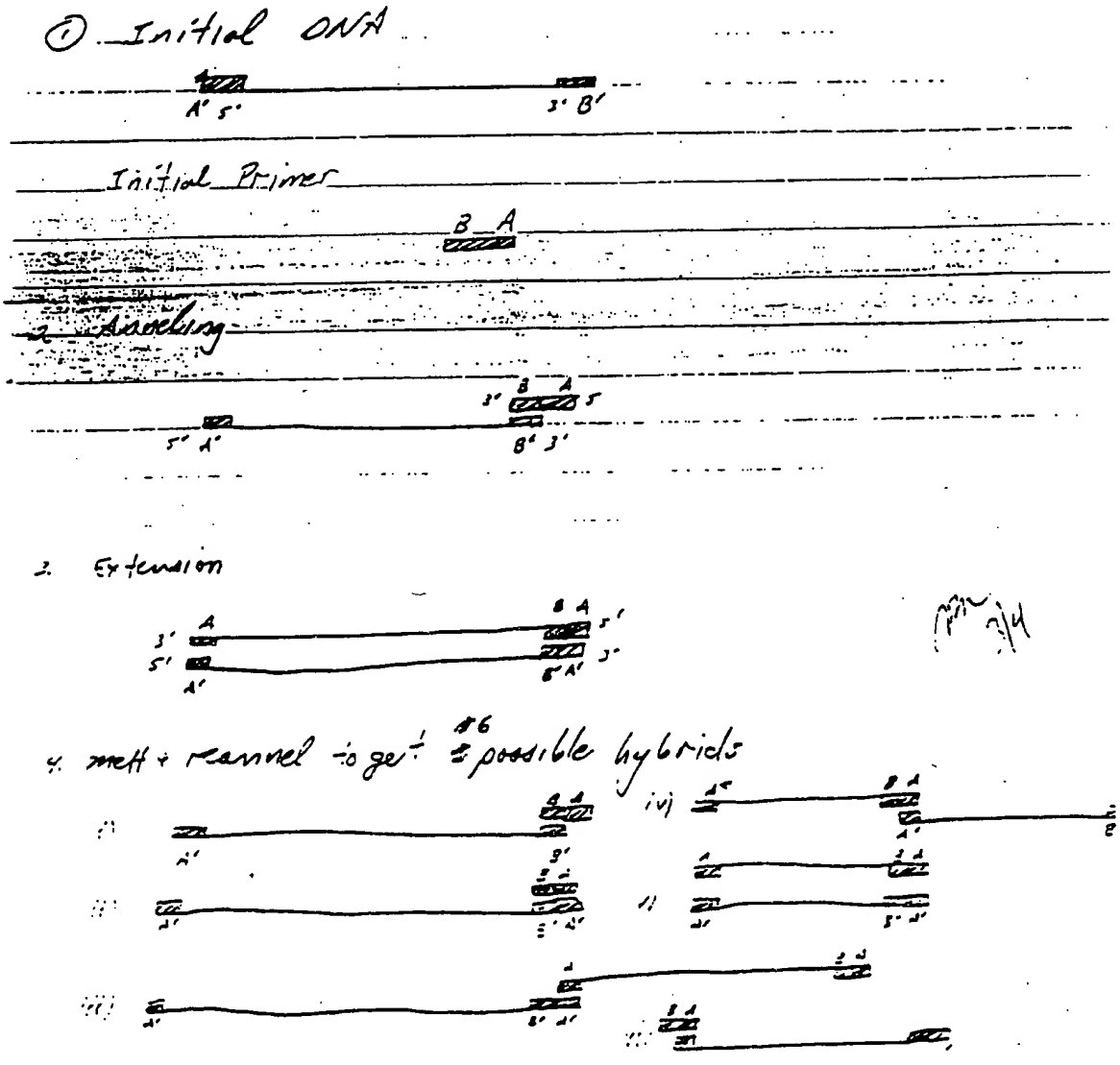
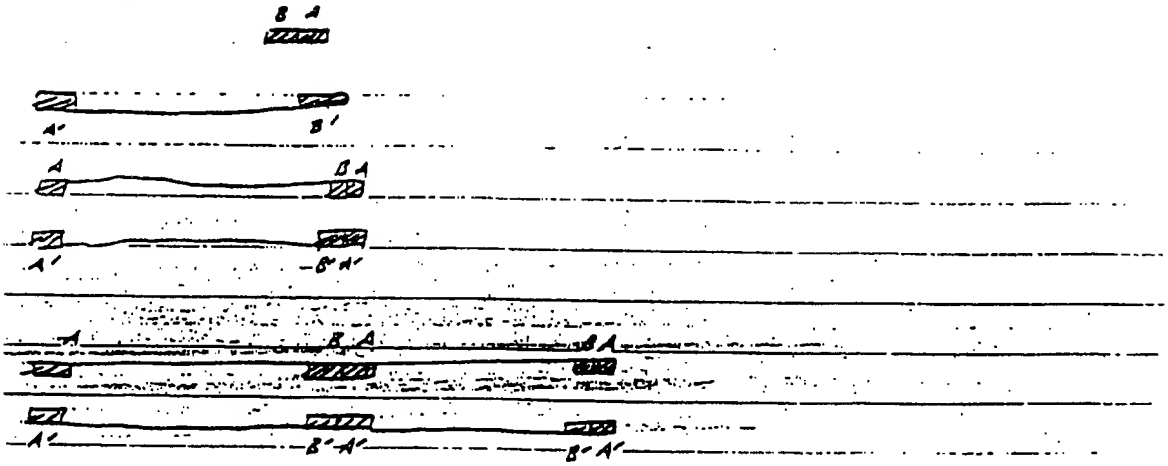


FIGURE 1B

Extend

only hybrids i, ii, iii extend.

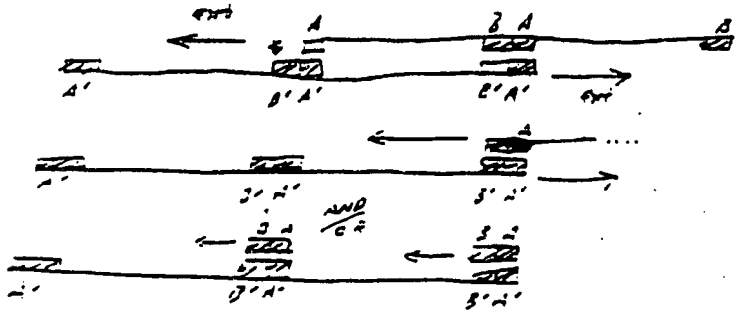
6. melt.



Reannealing gives the following possible extendible hybrids

i, ii, iii from step 4.

and.



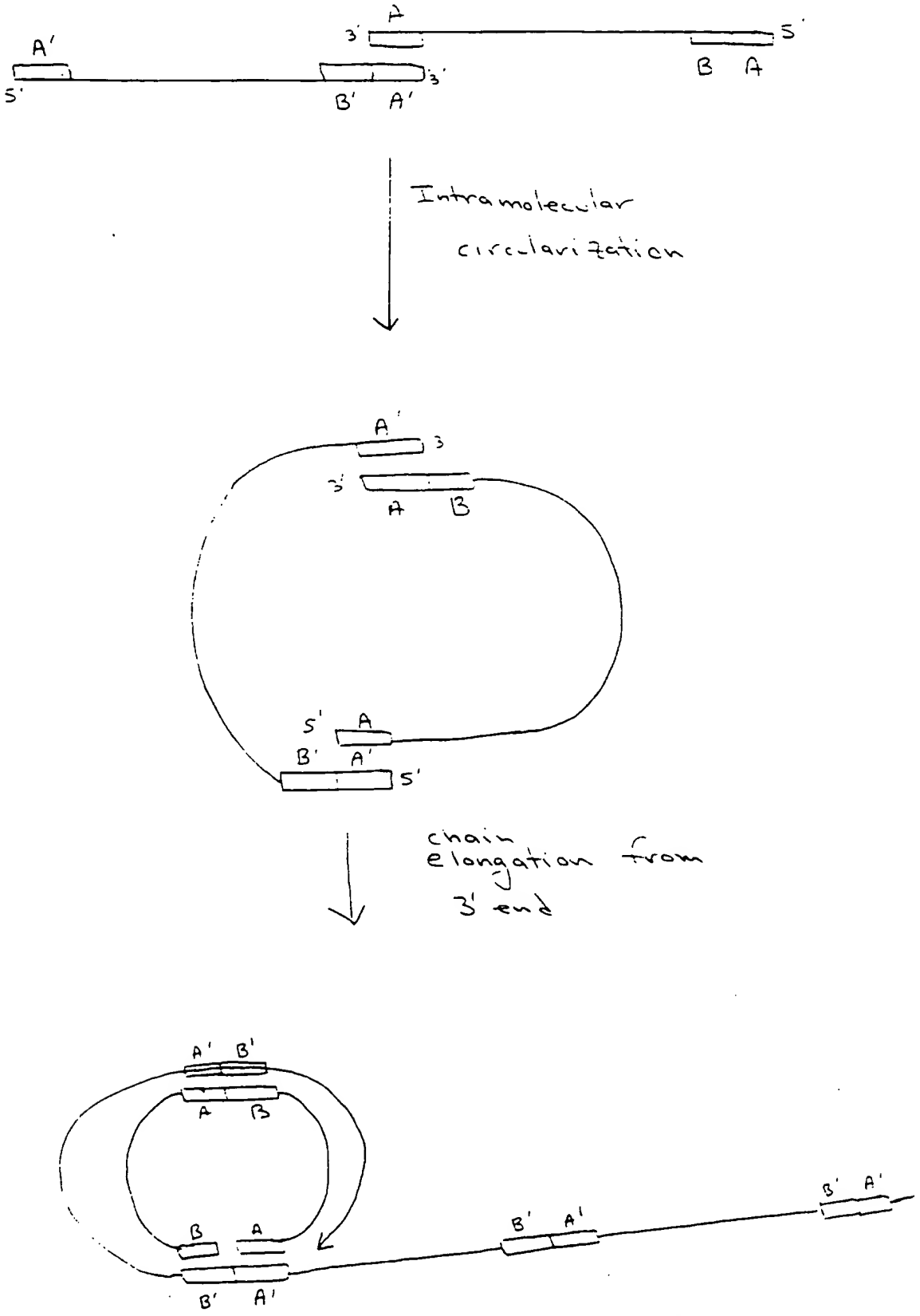

A'
5'
A
3'
5'
B A
B' A'
3'
Intramolecular
circularization
A'
3
3'
A B
5'
A
B'
A'
5'
chain
elongation from
3' end
A' B'
A B
B A
B' A'
B' A'
B' A'

FIGURE 2

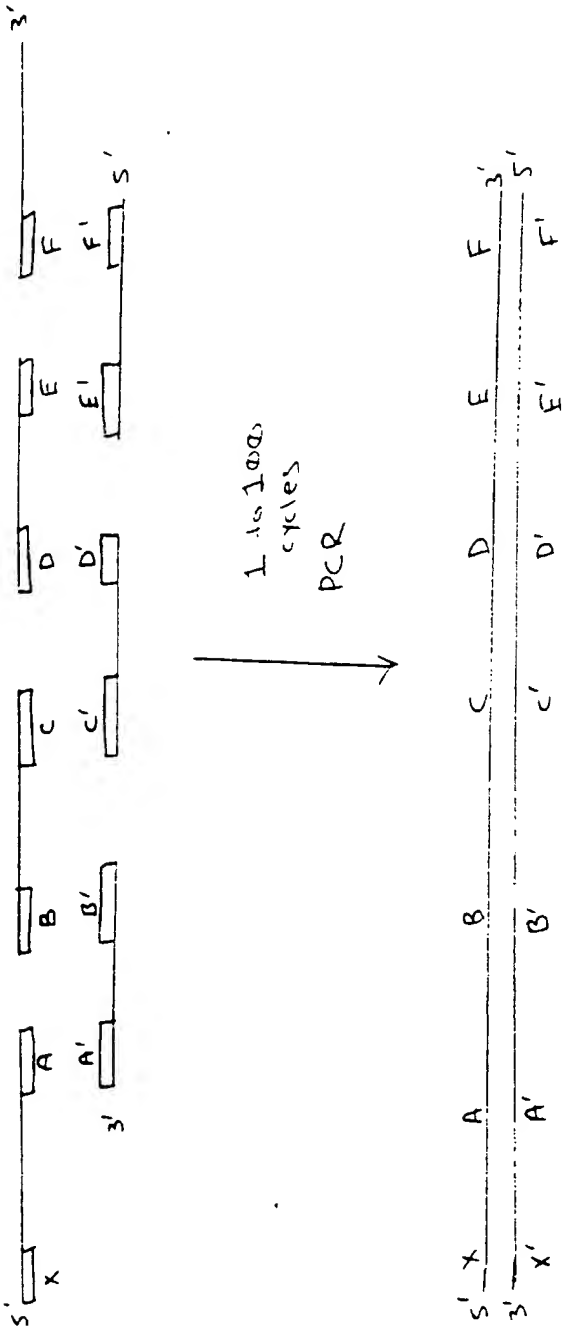


FIGURE 3

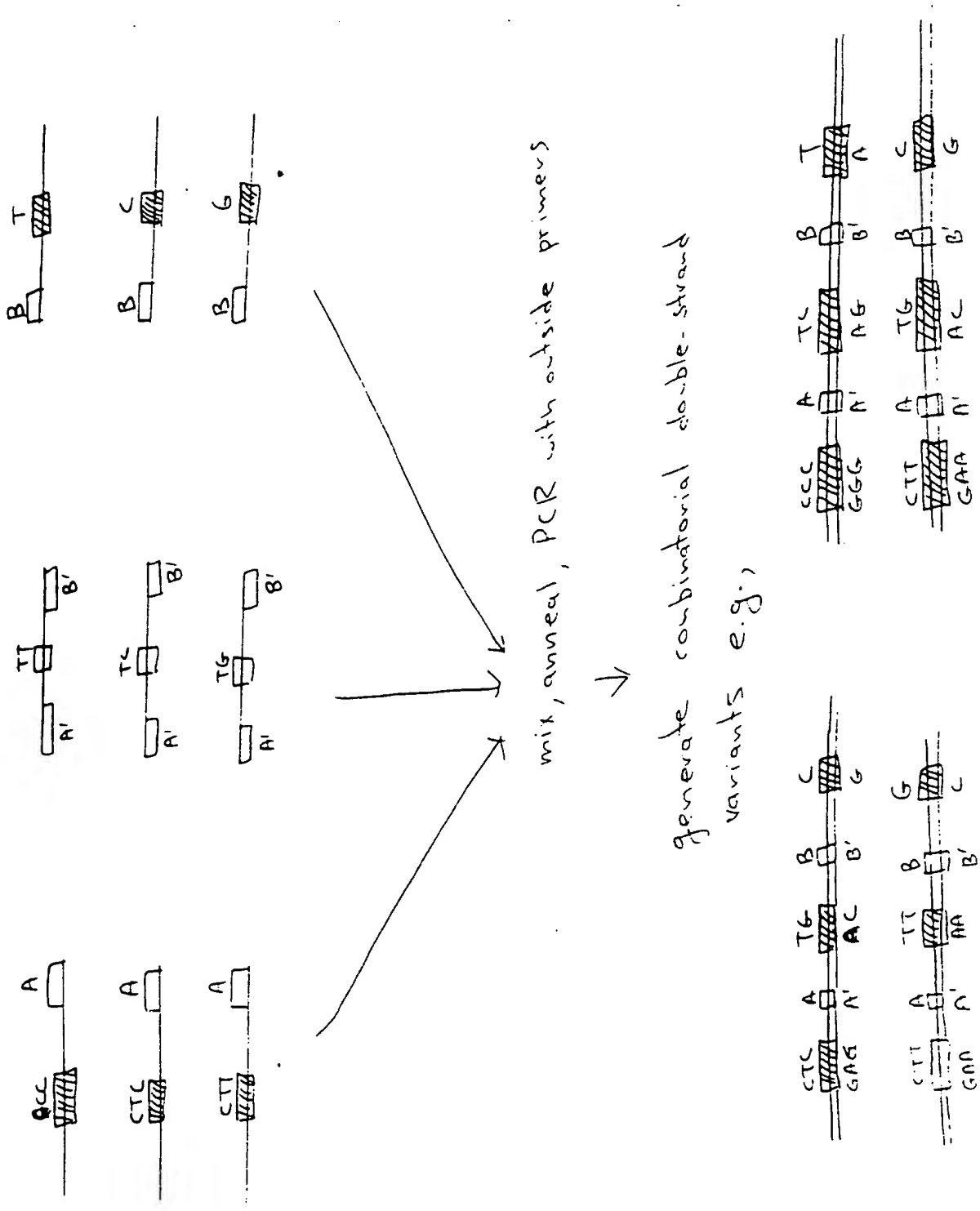

mix, anneal, PCR with outside primers
generate combinatorial double-strand variants e.g.,

FIGURE 4

R1 CTC ACG TTA AGG GAT TTT GGT CAT GAG ATT ATC AAA AAG G

R2 ATC TTC ACC TAG ATC CTT TTA AAT TAA AAA TGA AGT TTT A

R3 AAT CAA TCT AAA GTA TAT ATG AGG CCT GAC AGG CCG GTC T

R4 GAC AGT TAC CAA TGC TTA ATC AGT GAG GCA CCT ATC TCA G

R5 CGA TCT GTC TAT TTC GTT CAT CCA TGG TTG CCT GAC TCC C

R6 CGT CGT GTA GAT AAC TAC GAT ACG GGA GGG CTT ACC ATC T

R7 GGC CCC AGT GCT GCA ATG ATA CCG CGA GAC CCA CGC TCA C

R8 CGG CTC CAG ATT TAT CAG CAA TAA ACC AGC CAG CTG GAA G

R9 GGC CGA GCG CAG AAG TGG TCC TGC AAC TTT ATC CGC CTC C

R10 ATC CAG TCT ATT AAT TGT TGC CGG GAA GCT AGA GTA AGT A

R11 GTT CGC CAG TTA ATA GTT TGC GCA ACG TTG TTG CCA TGG C

R12 TAC AGG CAT CGT GGT GTC ACG CTC GTC GTT TGG AAT GGC T

R13 TCA TTC AGC TCC GGT TCC CAA CGA TCA AGG CGA GTT ACA T

R14 GAT CCC CCA TGT TGT GCA AAA AAG CGG TTA GCT CCT TCG G

R15 TCC TCC GAT CGT TGT CAG AAG TAA GTT GGC TGC AGT GTT A

R16 TCA CTC ATG GTT ATG GCA GCA CTG CAT AAT TCT CTT ACT G

R17 TCA TGC CAT CCG TAA GAT GCT TTT CTG TGA CTG GTG AGT A

R18 CTC AAC CAA GTC ATT CTG AGA ATA GTG TAT GCG GCG ACC G

R19 AGT TGC TCT TGC CCG GCG TCA ATA CGG GAT AAT ACC GCG C

R20 CAC ATA GCA GAA CTT TAA AAG TGC TCA TCA TTG GAA AAC G

R21 TTC TTC GGG GCG AAA ACT CTC AAG GAT CTT ACC GCT GTT G

R22 AGA TCC AGT TCG ATG TAA CCC ACT CGT GCA CCC AAC TGA T

R23 CTT CAG CAT CTT TTA CTT TCA CCA GCG TTT CTG GGT GAG C

R24 AAA AAC AGG AAG GCA AAA TGC CGC AAA AAA GGG AAT AAG G

R25 GCG ACA CGG AAA TGT TGA ATA CTC ATA CTC TTC CTT TTT C

R26 AAT ATT ATT GAA GCA TTT ATC AGT GTT ATT GTC TCA TGA G

R27 CGG ATA CAT ATT TGA ATG TAT TTA GGC CAT GGT GGC CAA A

R28 AAT AAA CAA ATA GGG GTT CCG CGC A[illegible] TTT CCC CGA AAA G

Fig 5A

```
1   TTC TTA GAC GTC AGG TGG CAC TTT TCG GGG AAA TGT GCG C
2   GGA ACC CCT ATT TGT TTA TTT TTG GCC ACC ATG GCC TAA A
3   TAC ATT CAA ATA TGT ATC CGC TCA TGA GAC AAT AAC CCT G
4   ATA AAT GCT TCA ATA ATA TTG AAA AAG GAA GAG TAT GAG T
5   ATT CAA CAT TTC CGT GTC GCC CTT ATT CCC TTT TTT GCG G
6   CAT TTT GCC TTC CTG TTT TTG CTC ACC CAG AAA CGC TGG T
7   GAA AGT AAA AGA TGC TGA AGA TCA GTT GGG TGC ACG AGT G
8   GGT TAC ATC GAA CTG GAT CTC AAC AGC GGT AAG ATC CTT G
9   AGA GTT TTC GCC CCG AAG AAC GTT TTC CAA TGA TGA GCA C
10  TTT TAA AGT TCT GCT ATG TGG CGC GGT ATT ATC CCG TAT T
11  GAC GCC GGG CAA GAG CAA CTC GGT CGC CGC ATA CAC TAT T
12  CTC AGA ATG ACT TGG TTG AGT ACT CAC CAG TCA CAG AAA A
13  GCA TCT TAC GGA TGG CAT GAC AGT AAG AGA ATT ATG CAG T
14  GCT GCC ATA ACC ATG AGT GAT AAC ACT GCA GCC AAC TTA C
15  TTC TGA CAA CGA TCG GAG GAC CGA AGG AGC TAA CCG CTT T
16  TTT GCA CAA CAT GGG GGA TCA TGT AAC TCG CCT TGA TCG T
17  TGG GAA CCG GAG CTG AAT GAA GCC ATT CCA AAC GAC GAG C
18  GTG ACA CCA CGA TGC CTG TAG CCA TGG CAA CAA CGT TGC G
19  CAA ACT ATT AAC TGG CGA ACT ACT TAC TCT AGC TTC CCG G
20  CAA CAA TTA ATA GAC TGG ATG GAG GCG GAT AAA GTT GCA G
21  GAC CAC TTC TGC GCT CGG CCC TTC CAG CTG GCT GGT TTA T
22  TGC TGA TAA ATC TGG AGC CGG TGA GCG TGG GTC TCG CGG T
23  ATC ATT GCA GCA CTG GGG CCA GAT GGT AAG CCC TCC CGT A
24  TCG TAG TTA TCT ACA CGA CGG GGA GTC AGG CAA CCA TGG A
25  TGA ACG AAA TAG ACA GAT CGC TGA GAT AGG TGC CTC ACT G
26  ATT AAG CAT TGG TAA CTG TCA GAC CGG CCT GTC AGG CCT C
27  ATA TAT ACT TTA GAT TGA TTT AAA ACT TCA TTT TTA ATT T
28  AAA AGG ATC TAG GTG AAG ATC CTT TTT GAT AAT CTC ATG A
```

Fig. 3B

(58F)
GGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCA
(59F)
CATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCC
(60F)
GCACCGATCGCCCTTCCCAACAGTTGCGTAGCCTGAATGG
(61F)
CGAATGGCGCCTGATGCGGTATTTTCTCCTTACGCATCTG
(62F)
TGCGGTATTTCACACCGCATATGGTGCACTCTCAGTACAA
(63F)
TCTGCTCTGATGCCGCATAGTTAAGCCAGCCCCGACACCC
(64F)
GCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTC
(65F)
CCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGA
(66F)
GCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACG
(30R)
CGAGGCCCTTTCGTCTCGCGCGTTTCGGTGATGACGGTGA
(31R)
AAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCT
(32R)
TGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGG
(33R)
GCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAA
(34R)
CTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCAT
(35R)
ATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAAT
(36R)
ACCGCATCAGGCGCCATTCGCCATTCAGGCTACGCAACTG
(37R)
TTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGC
(38R)
CAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTT
(39R)
GGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAAC
(40R)
GACGGCCAGTGCCAAGCTTGCATGCCTGCAGGTCGACTCT
(41R)
AGAGGATCCCCGGGTACCGAGCTCGAATTCGTAATCATGG
(42R)
TCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAA
(43R)
TTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGC
(44R)
CTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCG
(45R)
TTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGT
(46R)
GCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGG
(47R)
CGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACT
(48R)
GACTCGCTGCGCTCGGTCGTTCGGCTGCGCCGAGCGGTAT
(49R)
CAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATC

Fig. 5C

(50R)
AGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAG
(51R)
CAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGT
(52R)
TTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAA
(53R)
TCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTA
(54R)
TAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGC
(55R)
GCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTC
(56R)
CGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCAATGC
(57R)
TCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCT
(58R)
CCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGA
(59R)
CCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAAC
(60R)
CCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTG
(61R)
GTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTAC
(62R)
AGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGA
(63R)
AGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTA
(64R)
CCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACA
(65R)
AACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAG
(66R)
CAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTT
(67R)
TGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAA

(67F)
CGCGAGACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAA

(29R)
TGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTAT

Fig 5D

(29F)
CCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTC
(30F)
AGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCT
(31F)
TTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAAC
(32F)
CACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCT
(33F)
ACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCG
(34F)
CAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAG
(35F)
GCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCT
(36F)
CGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGC
(37F)
GATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGT
(38F)
TACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTC
(39F)
GTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAA
(40F)
CTGAGATACCTACAGCGTGAGCATTGAGAAAGCGCCACGC
(41F)
TTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGG
(42F)
CAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGG
(43F)
GGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCC
(44F)
ACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGG
(45F)
GGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTT
(46F)
TTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGT
(47F)
TCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTAT
(48F)
TACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGA
(49F)
ACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAG
(50F)
AGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCC
(51F)
GATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGG
(52F)
AAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGC
(53F)
TCACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCC
(54F)
GGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTT
(55F)
CACACAGGAAACAGCTATGACCATGATTACGAATTCGAGC
(56F)
TCGGTACCCGGGGATCCTCTAGAGTCGACCTGCAGGCATG
(57F)
CAAGCTTGGCACTGGCCGTCGTTTTACAACGTCGTGACTG

Fig. 2E

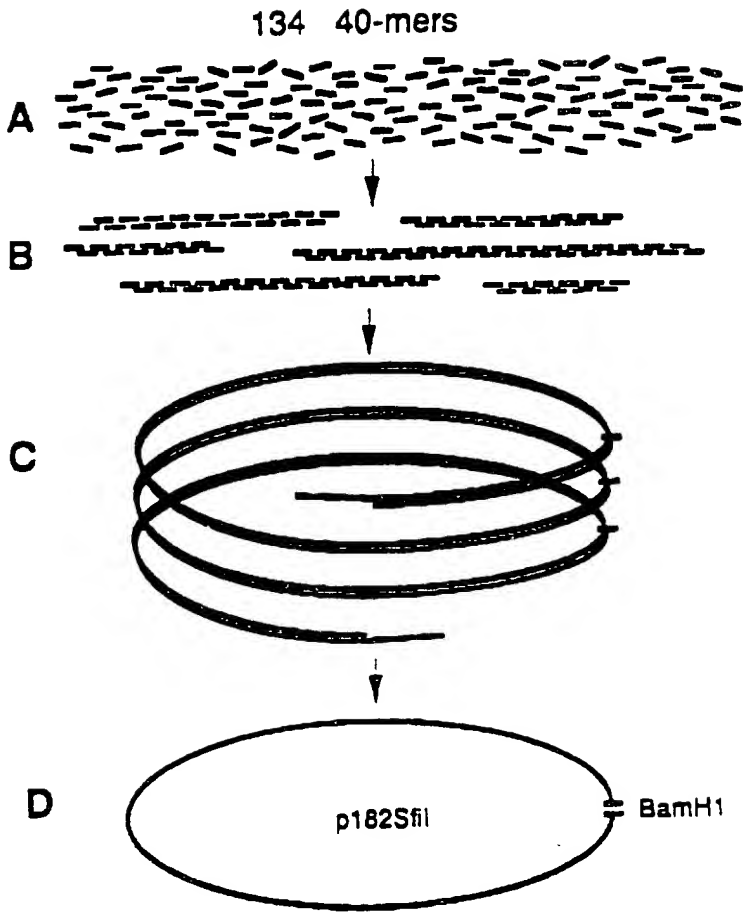


FIGURE 6

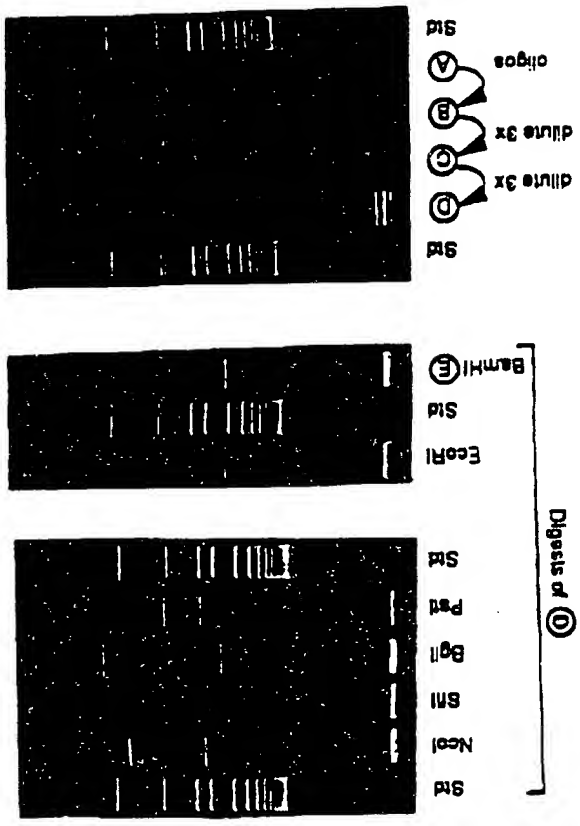

Std
oligos
A
dilute 3x
B
dilute 3x
C
Std
BamHI E
Std
EcoRI
Digests of D
Std
NcoI
SmI
BglI
PstI
Std

FIGURE 7

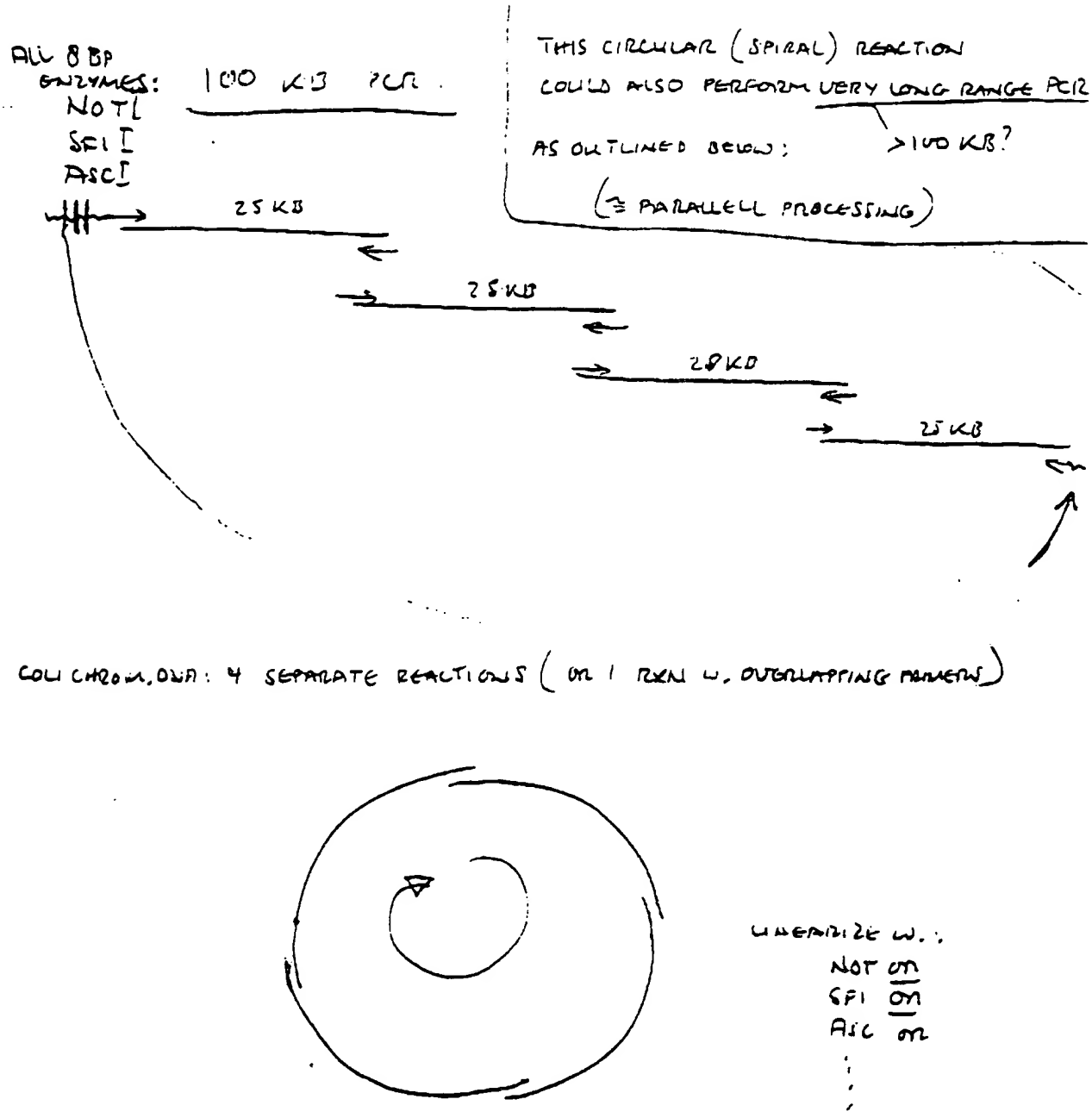
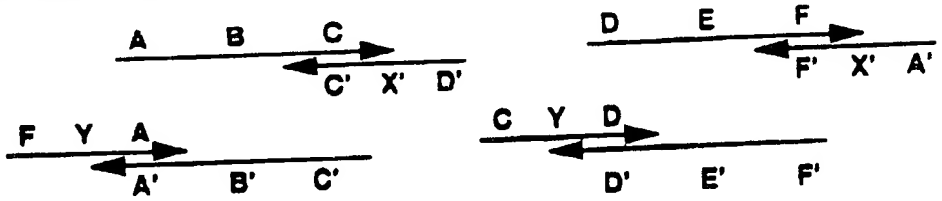

ALL 8 BP
ENZYMES:
NOT I
SFI I
ASC I
100 KB PCR
25 KB
25 KB
25 KB
25 KB
THIS CIRCULAR (SPIRAL) REACTION
COULD ALSO PERFORM VERY LONG RANGE PCR
AS OUTLINED BELOW;
>100 KB?
(= PARALLEL PROCESSING)
COLI CHROM. DNA: 4 SEPARATE REACTIONS (OR 1 RXN W. OVERLAPPING PRIMERS)
LINEARIZE W.:
NOT OR
SFI OR
ASC OR

Figure 8

Continuous Circular Multiplex PCR

Initial Conditions

Anneal with Primers



Extend and Melt

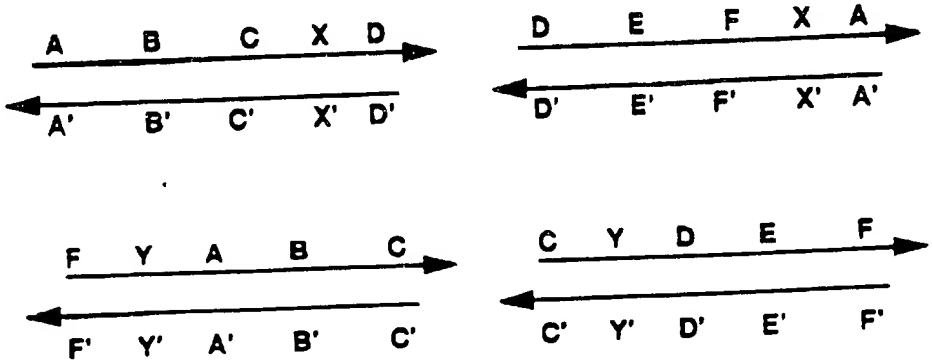


Figure 9A

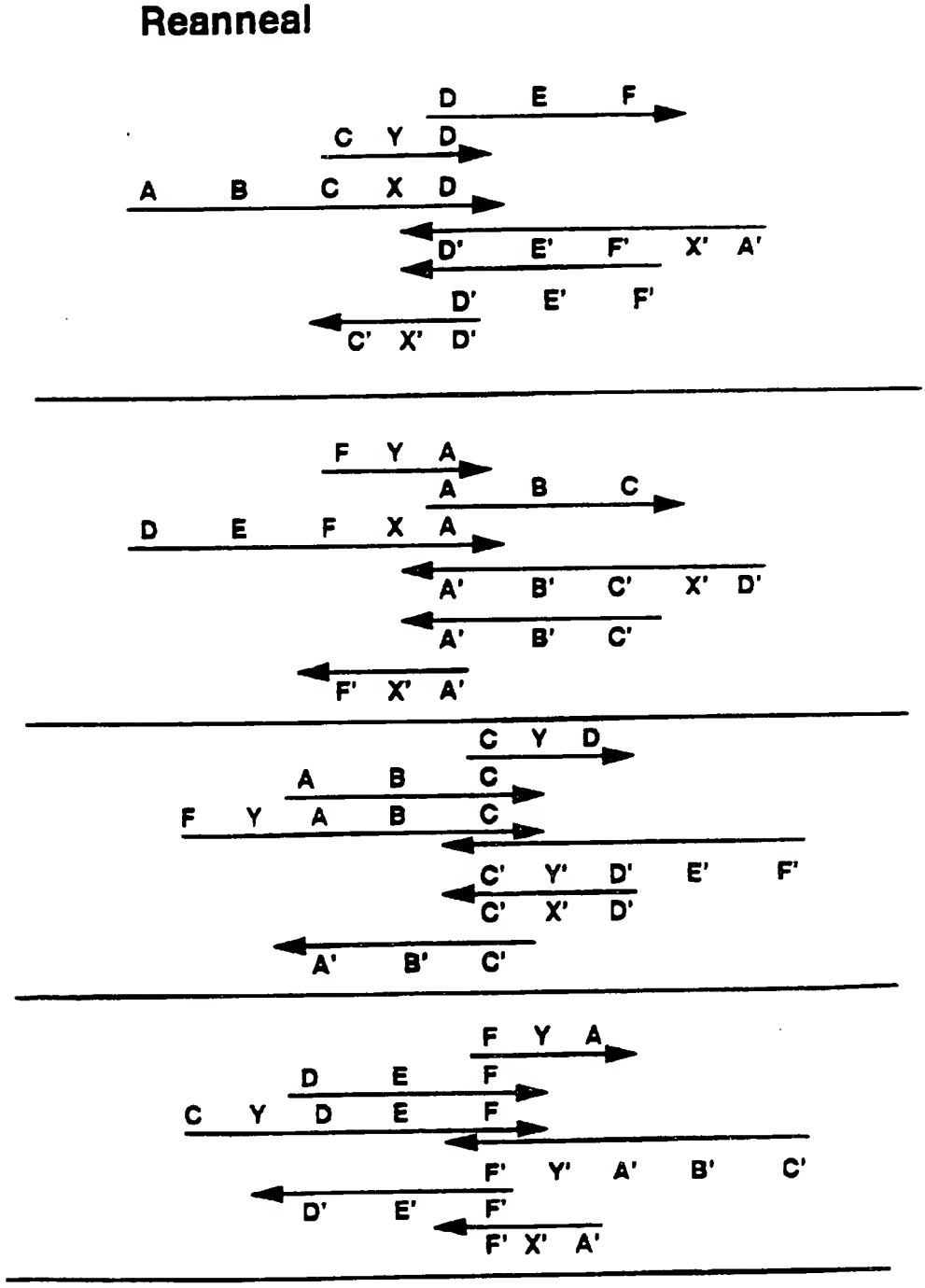

Reanneal
D E F
C Y D
A B C X D
D' E' F' X' A'
D' E' F'
C' X' D'
F Y A
A B C
D E F X A
A' B' C' X' D'
A' B' C'
F' X' A'
C Y D
A B C
F Y A B C
C' Y' D' E' F'
C' X' D'
A' B' C'
F Y A
D E F
C Y D E F
F' Y' A' B' C'
D' E' F'
F' X' A'

Figure 9B

Extend

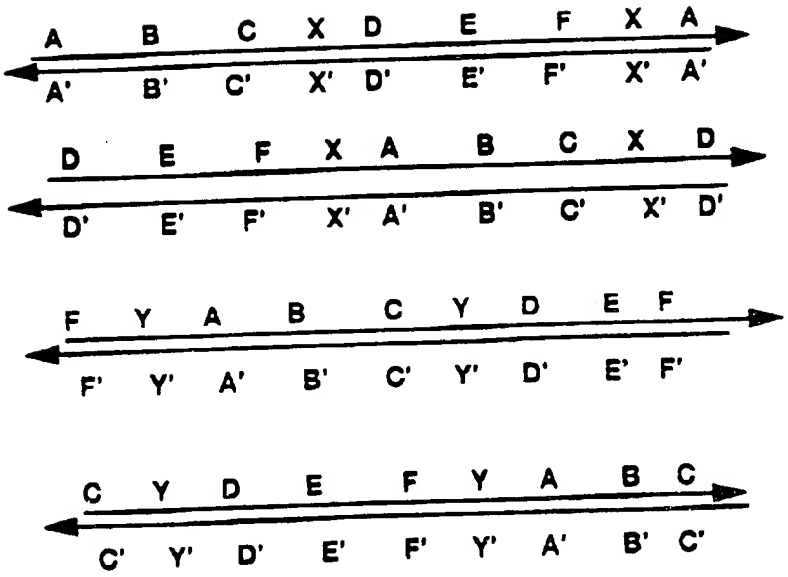
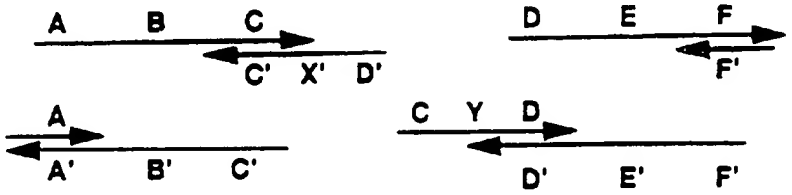


Figure 9C

Continuous Linear Multiplex PCR

Initial Conditions

Anneal with Primers



Extend and Melt

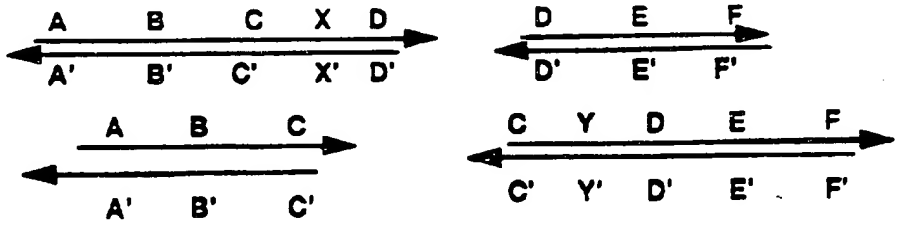


Figure 10A

Reann al

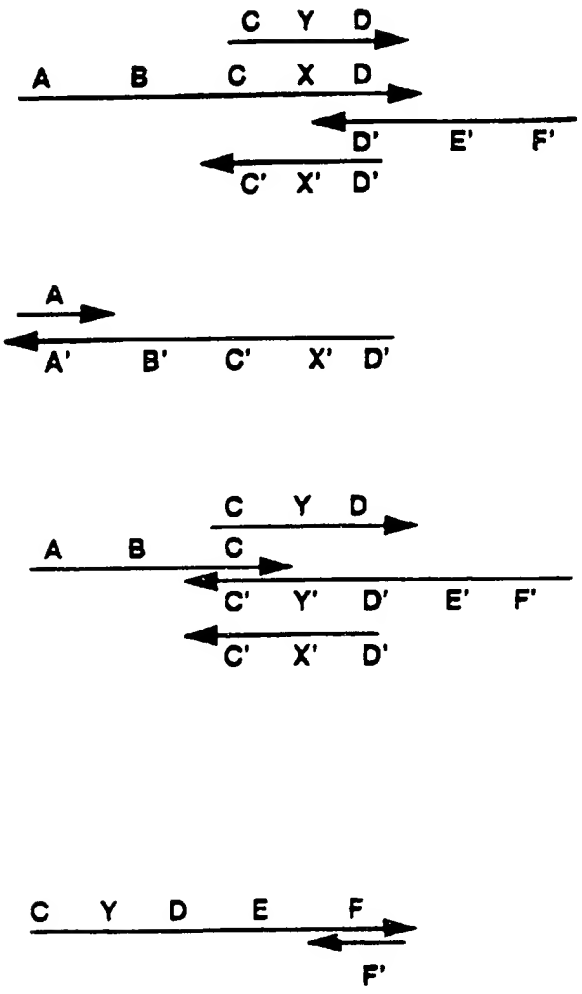
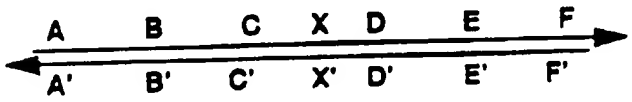
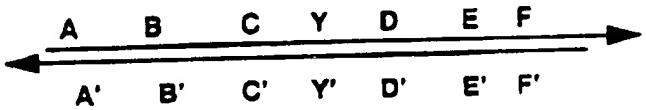


Figure 10B

Extend - After a period, assuming equimolar primers, only the following products are left.



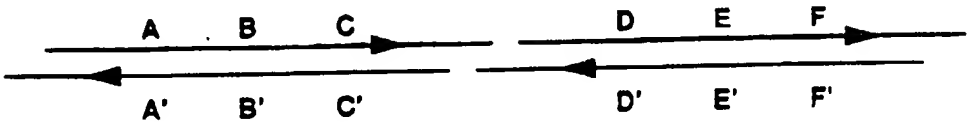


Amplify using only primers A and F '

If necessary, digest to separate the components using restriction sites incorporated into X and Y

Figure 10C

Initial Conditions: sequence embedded in genome DNA, indicated in gray



Anneal #1: Low concentration standard regular PCR Primers.

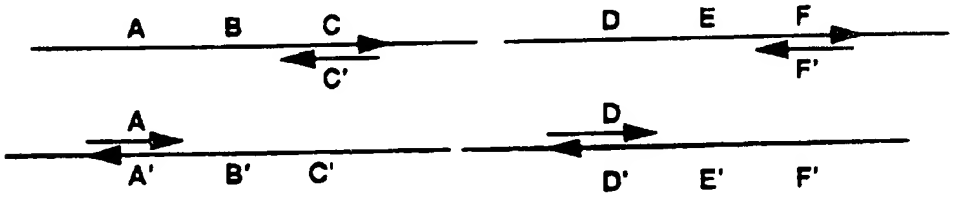
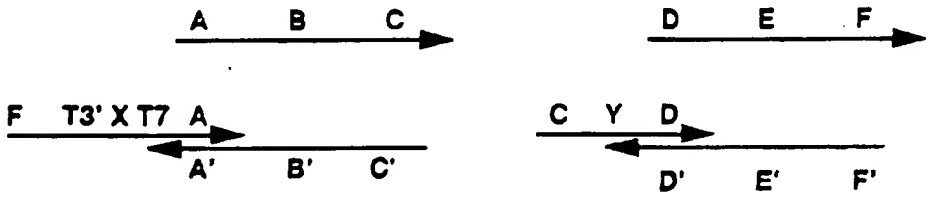


Figure 11A

Anneal with Primers: New primers are at higher concentration than initial primers so that there is no competition problem with these primers binding with the initial primers. Note that primers are selected for only one strand. The orientation is selected so that the resulting fragment overlap.



Extend and Melt:

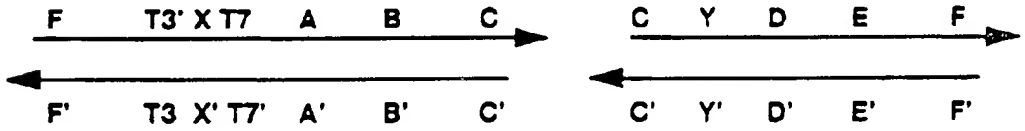
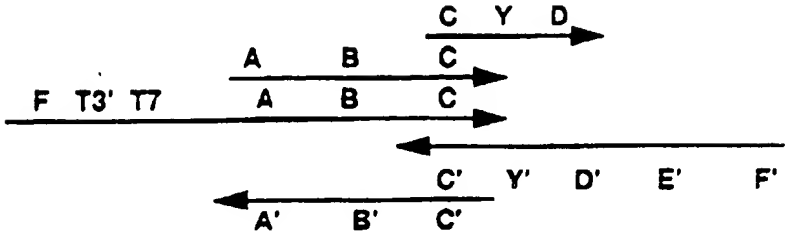
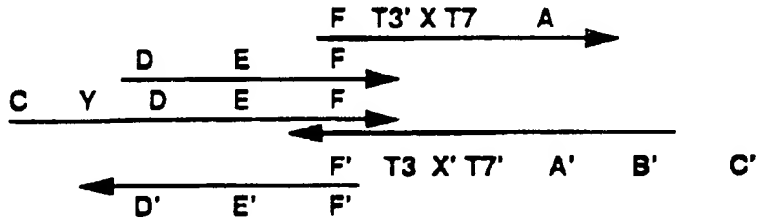


Figure 11 B

Reanneal





Extend

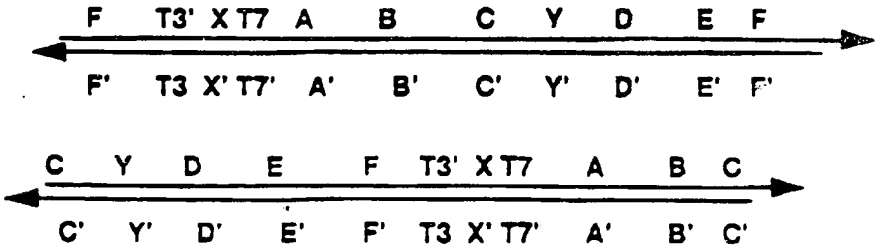


Figure 11C

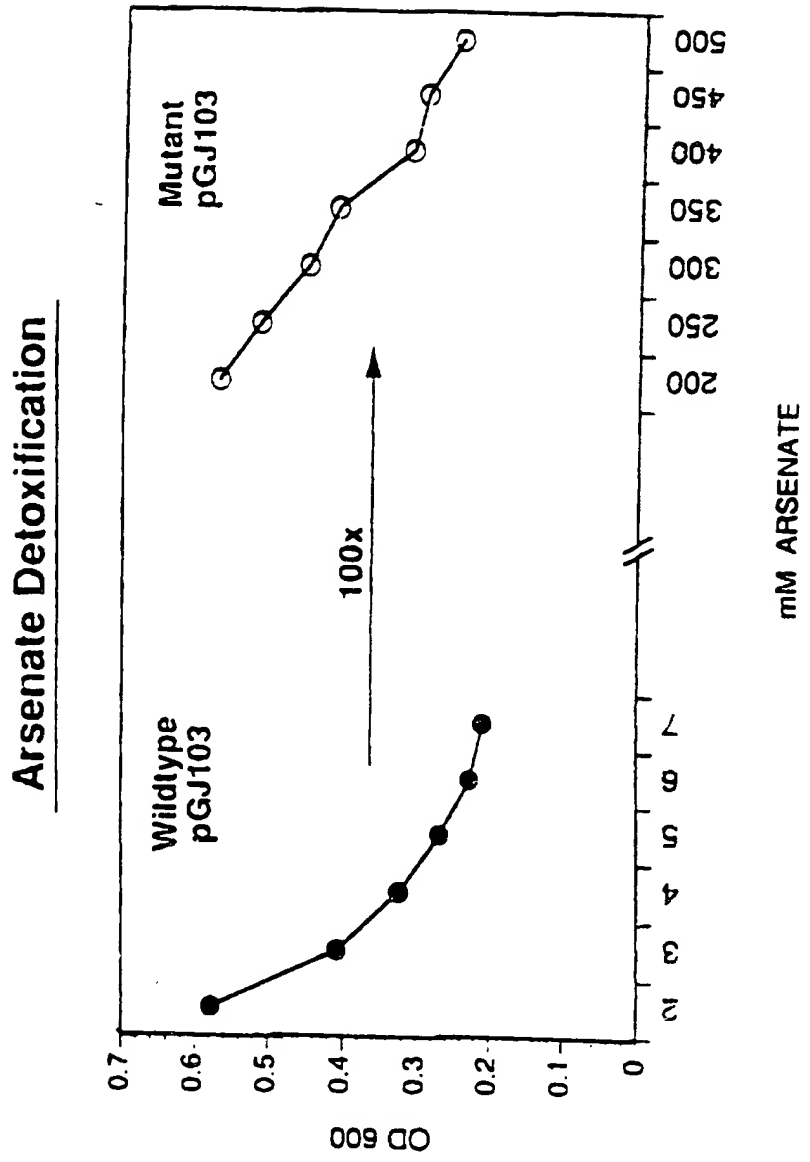

Arsenate Detoxification
Mutant pGJ103
Wildtype pGJ103
100x
OD 600
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
2
3
4
5
6
7
200
250
300
350
400
450
500
mM ARSENATE

Figure 12

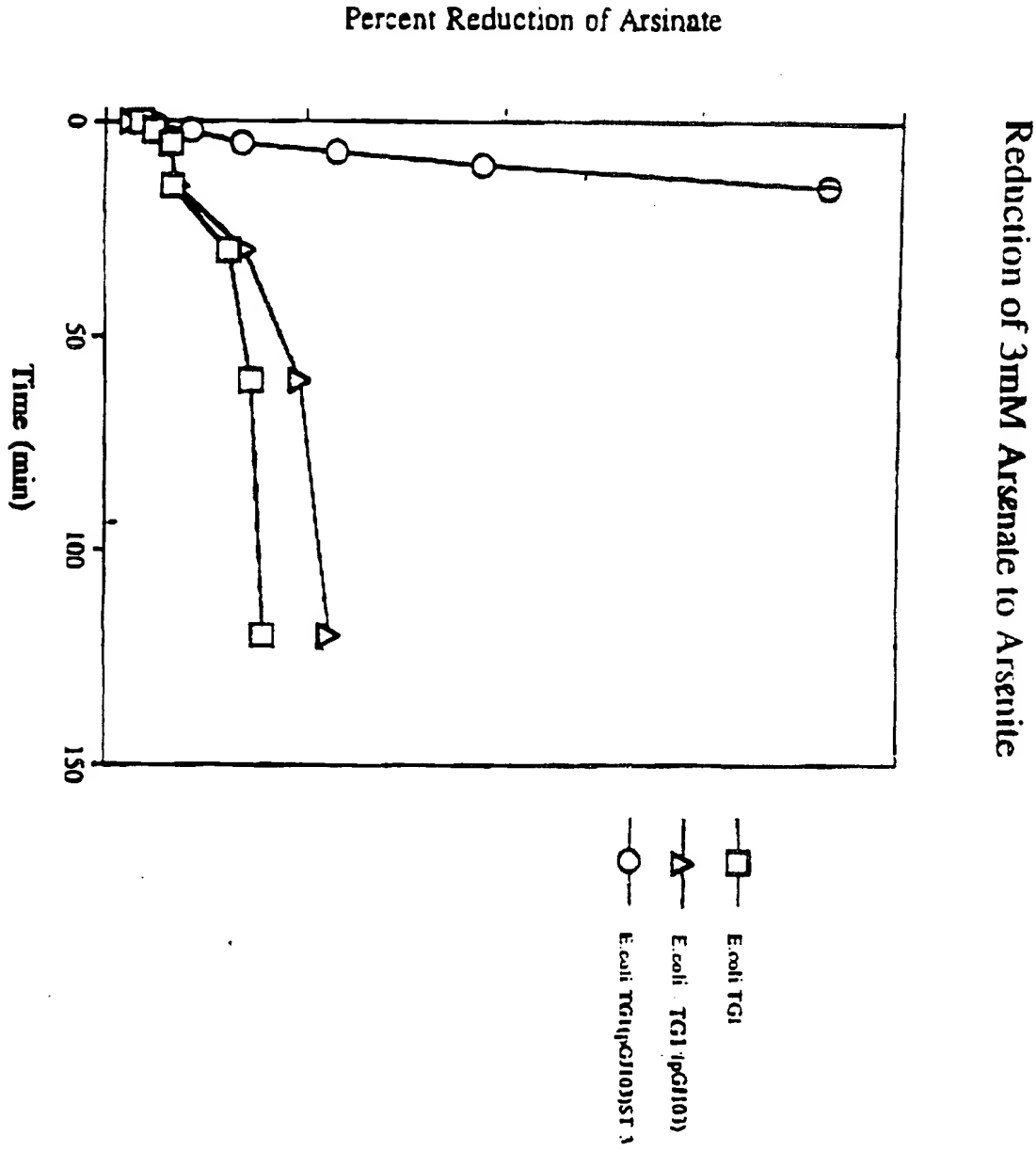

Percent Reduction of Arsinate
Time (min)
0
50
100
150
Reduction of 3mM Arsenate to Arsenite
E.coli TG1
E.coli TG1(pG110)
E.coli TG1(pG110)ST 1

Figure 13

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/05480

A. CLASSIFICATION OF SUBJECT MATTER
IPC(6) :C07H 21/04
US CL :536/24.3
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 536/24.3; 435/91.2

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

USPAT, MEDLINE, BIOSIS, CAPLUS
search terms: primers, PCR, head-tail , cycles, concatemer, ligation, amplification

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	WHITE et al., Analytical Biochemistry, "Concatemer Chain Reaction: A Tag DNA Polymerase-Mediated Mechanism for Generating Long Tandemly Repetitive DNA Sequences", Volume 199, issued 1991, pages 184-190, especially pages 186-187.	1, 2, 5-12, 14-18 ---------- 3, 4, 13
Y	US 5,118,604 A (WEISMAN et al) 02 June 1992 , see column 11, lines 30-50.	3
Y	SUZUKI et al. In vivo ligation of linear DNA molecules to circular forms in the yeast saccharomyces cerevisiae. Journal of Bacteriology. August 1983, Vol. 155, No. 2, pages 747-754, especially pages 747, 750, 752.	4, 13

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be part of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 22 JULY 1996	Date of mailing of the international search report 19 SEP 1996
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer Joyce Tung Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

From the INTERNATIONAL BUREAU

PCT

NOTIFICATION OF THE RECORDING OF A CHANGE

(PCT Rule 92bis.1 and Administrative Instructions, Section 422)

To:

MASCHIO, Antonio
D. Young & Co.
21 New Fetter Lane
London EC4A 1DA
ROYAUME-UNI

Date of mailing (day/month/year)
24 November 1999 (24.11.99)

Applicant's or agent's file reference
P004287WO ATM

IMPORTANT NOTIFICATION

International application No.
PCT/GB99/00816

International filing date (day/month/year)
17 March 1999 (17.03.99)

1. The following indications appeared on record concerning:

[X] the applicant [ ] the inventor [ ] the agent [ ] the common representative

Name and Address
MEDICAL RESEARCH COUNCIL
20 Park Crescent
London W1N 4AL
United Kingdom

State of Nationality: GB
State of Residence: GB
Telephone No.:
Facsimile No.:
Teleprinter No.:

2. The International Bureau hereby notifies the applicant that the following change has been recorded concerning:

[X] the person [ ] the name [ ] the address [ ] the nationality [ ] the residence

Name and Address
GENDAQ LIMITED
1-3 Burtonhole Lane
Mill Hill
London NW7 1AD
United Kingdom

State of Nationality: GB
State of Residence: GB
Telephone No.:
Facsimile No.:
Teleprinter No.:

3. Further observations, if necessary:

4. A copy of this notification has been sent to:

[X] the receiving Office
[ ] the International Searching Authority
[X] the International Preliminary Examining Authority
[ ] the designated Offices concerned
[X] the elected Offices concerned
[ ] other:

The International Bureau of WIPO
34, chemin des Colombettes
1211 Geneva 20, Switzerland

Facsimile No.: (41-22) 740.14.35

Authorized officer
H. Zhou

Telephone No.: (41-22) 338.83.38

Form PCT/IB/306 (March 1994)

002970354

From the INTERNATIONAL BUREAU

PCT

NOTIFICATION OF ELECTION

(PCT Rule 61.2)

To:

Assistant Commissioner for Patents
United States Patent and Trademark Office
Box PCT
Washington, D.C.20231
ÉTATS-UNIS D'AMÉRIQUE

in its capacity as elected Office

Date of mailing (day/month/year) 24 November 1999 (24.11.99)	
International application No. PCT/GB99/00816	**Applicant's or agent's file reference** P004287WO ATM
International filing date (day/month/year) 17 March 1999 (17.03.99)	**Priority date** (day/month/year) 17 March 1998 (17.03.98)
Applicant CHOO, Yen et al	

1. The designated Office is hereby notified of its election made:

[X] in the demand filed with the International Preliminary Examining Authority on:

06 October 1999 (06.10.99)

[ ] in a notice effecting later election filed with the International Bureau on:

2. The election [X] was

[ ] was not

made before the expiration of 19 months from the priority date or, where Rule 32 applies, within the time limit under Rule 32.2(b).

The International Bureau of WIPO
34, chemin des Colombettes
1211 Geneva 20, Switzerland

Facsimile No.: (41-22) 740.14.35

Authorized officer

H. Zhou

Telephone No.: (41-22) 338.83.38

PATENT COOPERATION TREATY

PTO/PCT Rec'd 18 SEP 2000

From the INTERNATIONAL BUREAU

PCT

NOTIFICATION OF THE RECORDING OF A CHANGE

(PCT Rule 92bis.1 and Administrative Instructions, Section 422)

To:

MASCHIO, Antonio
D. Young & Co.
21 New Fetter Lane
London EC4A 1DA
ROYAUME-UNI

Records noted

Date of mailing (day/month/year)
24 November 1999 (24.11.99)

Applicant's or agent's file reference
P004287WO ATM

IMPORTANT NOTIFICATION

International application No.
PCT/GB99/00816

International filing date (day/month/year)
17 March 1999 (17.03.99)

1. The following indications appeared on record concerning:

[X] the applicant [ ] the inventor [ ] the agent [ ] the common representative

Name and Address
MEDICAL RESEARCH COUNCIL
20 Park Crescent
London W1N 4AL
United Kingdom

State of Nationality: GB
State of Residence: GB
Telephone No.:
Facsimile No.:
Teleprinter No.:

2. The International Bureau hereby notifies the applicant that the following change has been recorded concerning:

[X] the person [ ] the name [ ] the address [ ] the nationality [ ] the residence

Name and Address
GENDAQ LIMITED
1-3 Burtonhole Lane
Mill Hill
London NW7 1AD
United Kingdom

State of Nationality: GB
State of Residence: GB
Telephone No.:
Facsimile No.:
Teleprinter No.:

3. Further observations, if necessary:

4. A copy of this notification has been sent to:

[X] the receiving Office
[ ] the International Searching Authority
[X] the International Preliminary Examining Authority
[ ] the designated Offices concerned
[X] the elected Offices concerned
[ ] other:

The International Bureau of WIPO
34, chemin des Colombettes
1211 Geneva 20, Switzerland

Facsimile No.: (41-22) 740.14.35

Authorized officer
H. Zhou

Telephone No.: (41-22) 338.83.38

PATENT ASSIGNMENT

DATE: 21 MAY 1999

PARTIES:

(1) MEDICAL RESEARCH COUNCIL (a body incorporated by Royal Charter) of 20 Park Crescent, London W1N 4AL, England ("**Assignor**")

(2) ENDLOCK LIMITED (Registered in England No. 3756817) whose registered office is at 90 Long Acre, London WC2E 9TT, England ("**Assignee**")

RECITALS:

(A) The ASSIGNOR is the applicant in respect of the patent applications named in the Schedule to this Assignment (the "**Applications**") in respect of the inventions disclosed in the Applications (the "**Inventions**").

(B) The parties have agreed that all the Assignor's rights in the Applications and the Inventions shall be assigned to the Assignee.

OPERATIVE TERMS:

In consideration of the sum of £1,450 paid by the Assignee to the Assignor (the receipt and sufficiency of which is hereby acknowledged) IT IS AGREED as follows:

1. Assignment

The Assignor hereby assigns to the Assignee throughout the world:

(a) the Inventions and all right and title of the Assignor therein;

(b) the Applications and all patent applications, patents and supplementary protection certificates which may derive or result from any of the same, including any divisionals, continuations, reissues, extensions and registrations in relation to any thereof and all and any other rights in the inventions comprised in the above (together referred to as the "**Patent Rights**");

(c) the benefit of all priority dates;

(d) its right to apply for and obtain patents, registrations or similar forms of protection in other countries in respect of the Inventions or otherwise included in or derived from the Patent Rights;

(e) its right to make any new application(s) in respect of any parts of the subject matter of any application or specification filed in connection with the Inventions and the right to claim priority from the Applications;

(f) its right to sue for and obtain injunctive relief, damages and all other relief in respect of any infringement or misuse (whether past present or future) of the Applications or of the rights conferred by publication of the same or of any other of the Patent Rights.

2. Validity

The Assignor has made available to the Assignee its case information concerning the current state and progress of the Applications but nothing in this Assignment shall be construed as a representation or warranty that any of the Applications will proceed to grant or that any of the Patent Rights are valid or that any exploitation of the Inventions represented by the Applications will not be an infringement of any patents or other rights not vested in the Assignor.

3. Further Assurance

The Assignor agrees at the request and expense of the Assignee to execute any further documents and do all things which the Assignee may reasonably require in order to vest in the Assignee (or the Assignee's successors in title or nominees as the case may be) the Patent Rights including the Applications and all other rights title and interests intended to be assigned transferred or granted to the Assignee hereunder and to give the Assignee the full benefit of this Assignment including doing all acts which may be necessary or desirable to record the Applications and any other Patent Rights in the name of the Assignee in any relevant register of patent rights.

4. Law

This Assignment shall be governed by and construed and take effect in accordance with English law.

IN WITNESS of which the parties have executed this Assignment on the date shown at the beginning of this Assignment.

SCHEDULE

Number	
PCT/GB95/01949	(granted in Australia, but pending elsewhere)
PCT/GB98/01510	
PCT/GB98/01512	
PCT/GB98/01514	
PCT/GB98/01516	
GB:980 5576.7	
GB:980 6895.0	
GB: 980 7246.5	

GB: 9824544.2 (originally GB: 9723430.6)

EXECUTED for and on behalf of
MEDICAL RESEARCH COUNCIL

Signature: .....

Name: Graham L. Wagner
Licensing and Agreements Manager

Title: .....

EXECUTED for and on behalf of
ENDLOCK LIMITED

Signature: .....

Name: Stephen T. Reed

Title: Director

CERTIFICATE OF INCORPORATION

ON CHANGE OF NAME

Company No. 3756817

The Registrar of Companies for England and Wales hereby certifies that

ENDLOCK LIMITED

having by special resolution changed its name, is now incorporated

under the name of

GENDAQ LIMITED

Given at Companies House, Cardiff, the 24th June 1999

REGISTRAR OF COMPANIES FOR ENGLAND AND WALES

THE FFICIAL SEAL F THE REGISTRAR F COMPANIES

COMPANIES HOUSE

HC006A

PTO/PCT 18 SEP 2000

PCT

REQUEST

The undersigned requests that the present international application be processed according to the Patent Cooperation Treaty.

For receiving office use only

International Application No.

International Filing Date

Name of receiving Office and "PCT International Application"

Applicant's or agent's file reference *(if desired) (12 characters maximum)* P004287WO ATM

Box No. I TITLE OF INVENTION NUCLEIC ACID BINDING PROTEINS

Box No. II APPLICANT

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (i.e. country) of residence if no State of residence is indicated below.)*

Medical Research Council
20 Park Crescent
London
W1N 4AL
United Kingdom

☐ This person is also inventor.

Telephone No.

Facsimile No.

Teleprinter No.

State *(i.e. country)* of nationality: United Kingdom

State *(i.e. country)* of residence: United Kingdom

This person is applicant for the purposes of: ☐ all designated States ☑ all designated States except the United States of America ☐ the United States of America only ☐ the States indicated in the Supplemental Box

Box No. III FURTHER APPLICANT(S) AND/OR (FURTHER) INVENTOR(S)

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (i.e. country) of residence if no State of residence is indicated below.)*

CHOO, Yen
MRC Laboratory of Molecuar Biology
Medical Research Council
Hills Road
Cambridge
CB2 2QH United Kingdom

This person is:

☐ applicant only

☑ applicant and inventor

☐ inventor only *(if this check-box is marked, do not fill in below)*

State *(i.e. country)* of nationality: Greece

State *(i.e. country)* of residence: United Kingdom

This person is applicant for the purposes of: ☐ all designated States ☐ all designated States except the United States of America ☑ the United States of America only ☐ the States indicated in the Supplemental Box

☑ Further applicant and/or (further) inventors are indicated on a continuation sheet

Box No. IV AGENT OR COMMON REPRESENTATIVE; OR ADDRESS FOR CORRESPONDENCE

The person identified below is hereby/has been appointed to act on behalf of the applicant(s) before the competent International Authorities as: ☑ agent ☐ common representative

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country.)*

MASCHIO, Antonio
D Young & Co
21 New Fetter Lane
London
EC4A 1DA
United Kingdom

Telephone No. +44 1703 634816

Facsimile No. +44 1703 224262

Teleprinter No. 477667 YOUNGS G

☐ Mark this check-box where no agent or common representative is/has been appointed and the space above is used instead to indicate a special address to which correspondence should be sent.

Continuation of Box No. III FURTHER APPLICANTS AND/OR (FURTHER) INVENTORS

If none of the following sub-boxes is used, this sheet is not to be included in the request.

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (i.e. country) of residence if no State of residence is indicated below.)*

ISALAN, Mark
24 Shottfield Avenue
East Sheen
London
SW14 8EA
United Kingdom

This person is:

- ☐ applicant only
- ☑ applicant and inventor
- ☐ inventor only *(if this check-box is marked, do not fill in below)*

State *(i.e. country)* of nationality: United Kingdom

State *(i.e. country)* of residence: United Kingdom

This person is applicant for the purposes of: ☐ all designated States ☐ all designated States except the United States of America ☑ the United States of America only ☐ the States indicated in the Supplemental Box

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (i.e. country) of residence if no State of residence is indicated below.)*

This person is:

- ☐ applicant only
- ☐ applicant and inventor
- ☐ inventor only *(if this check-box is marked, do not fill in below)*

State *(i.e. country)* of nationality:

State *(i.e. country)* of residence:

This person is applicant for the purposes of: ☐ all designated States ☐ all designated States except the United States of America ☐ the United States of America only ☐ the States indicated in the Supplemental Box

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (i.e. country) of residence if no State of residence is indicated below.)*

This person is:

- ☐ applicant only
- ☐ applicant and inventor
- ☐ inventor only *(if this check-box is marked, do not fill in below)*

State *(i.e. country)* of nationality:

State *(i.e. country)* of residence:

This person is applicant for the purposes of: ☐ all designated States ☐ all designated States except the United States of America ☐ the United States of America only ☐ the States indicated in the Supplemental Box

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (i.e. country) of residence if no State of residence is indicated below.)*

This person is:

- ☐ applicant only
- ☐ applicant and inventor
- ☐ inventor only *(if this check-box is marked, do not fill in below)*

State *(i.e. country)* of nationality:

State *(i.e. country)* of residence:

This person is applicant for the purposes of: ☐ all designated States ☐ all designated States except the United States of America ☐ the United States of America only ☐ the States indicated in the Supplemental Box

☐ Further applicants and/or (further) inventors are indicated on a continuation sheet

Supplemental Box

If the Supplemental Box is not used, this sheet need not be included in the request.

Use this box in the following cases:

1. If, in any of the Boxes, the space is insufficient to furnish all the information:

in particular:

(i) if more than two persons are involved as applicants and/or inventors and no "continuation sheet" is available:

(ii) if, in Box No. II or in any of the sub-boxes of Box No. III, the indication "the States indicated in the Supplemental Box" is checked:

(iii) if, in Box No. II or in any of the sub-boxes of Box No. III, the inventor or the inventor/applicant is not inventor for the purposes of all designated States or for the purposes of the United States of America:

(iv) if, in addition to the agent(s) indicated in Box No. IV, there are further agents:

(v) if, in Box No. V, the name of any State (or OAPI) is accompanied by the indication "patent of addition," or "certificate of addition," or if, in Box No. V, the name of the United States f America is accompanied by an indication "Continuation" or "Continuation-in-part":

(vi) if there are more than three earlier applications whose priority is claimed:

in such case, write "Continuation of Box No. ..." [indicate the number of the Box] and furnish the information in the same manner as required according to the captions of the Box in which the space was insufficient;

in such case, write "Continuation of Box No. III" and indicate for each additional person the same type of information as required in Box No. III. The country of the address indicated in this Box is the applicant's State (i.e. country) of residence if no State of residence is indicated below;

in such case, write "Continuation of Box No. II" or "Continuation of Box No. III" or "Continuation of Boxes No. II and III" (as the case may be), indicate the name of the applicant(s) involved and, next to (each) such name, State(s) (and/or, where applicable, ARIPO, Eurasian, European or OAPI patent) for the purposes of which the named person is applicant;

in such case, write "Continuation of Box No. II" or "Continuation of Box No. III" or "Continuation of Boxes No. II and III" (as the case may be), indicate the name of the inventor(s) and, next to (each) such name, State(s) (and/or, where applicable, ARIPO, Eurasian, European or OAPI patent) for the purposes of which the named person is inventor;

in such case, write "Continuation of Box No. IV and indicate for each further agent the same type of information as required in Box No. IV;

in such case, write "Continuation of Box No. V" and the name of each State involved (or OAPI), and after the name of each such State (or OAPI), the number of the parent title or parent application and the date of grant of the parent title or filing of the parent application;

in such case, write "Continuation of Box No. VI" and indicate for each additional earlier application the same type of information as required in Box No. VI.

2. If the applicant claims, in respect of any designated Office, the benefits of provisions of the national law concerning non-prejudicial disclosures or exceptions to lack of novelty:

in such case, write "Statement Concerning Non-Prejudicial Disclosures or Exceptions to Lack of Novelty" and furnish that statement below.

CONTINUATION OF BOX IV - ADDITIONAL REPRESENTATIVES

PURVIS, William Michael Cameron
COTTER, Ivan John
PILCH, Adam Johyn Michael
CRISP, David Norman
ROBINSON, Nigel Alexander Julian
HARRIS, Ian Richard
TURNER, James Arthur
HARDING, Charles Thomas
MALLALIEU, Catherine Louise
PRICE, Paul Anthony King
PRATT, Richard Wilson
HOLMES, Miles Keeton
HORNER, David
NACHSHEN, Neil
POTTER, Julian Mark

Box No. V DESIGNATION OF STATES

The following designations are hereby made under Rule 4.9(a) *(mark the applicable check-boxes; at least one must be marked)*:

Regional Patent

- ☑ **AP** **ARIPO Patent:** **GH** Ghana, **GM** Gambia, **KE** Kenya, **LS** Lesotho, **MW** Malawi, **SD** Sudan, **SZ** Swaziland, **UG** Uganda, **ZW** Zimbabwe, and any other State which is a Contracting State of the Harare Protocol and of the PCT
- ☑ **EA** **Eurasian Patent:** **AM** Armenia, **AZ** Azerbaijan, **BY** Belarus, **KG** Kyrgyzstan, **KZ** Kazakstan, **MD** Republic of Moldova, **RU** Russian Federation, **TJ** Tajikstan, **TM** Turkmenistan, and any other State which is a Contracting State of the Eurasian Patent Convention and of the PCT
- ☑ **EP** **European Patent:** **AT** Austria, **BE** Belgium, **CH and LI** Switzerland and Liechtenstein,**CY** Cyprus, **DE** Germany, **DK** Denmark, **ES** Spain, **FI** Finland, **FR** France, **GB** United Kingdom, **GR** Greece, **IE** Ireland, **IT** Italy, **LU** Luxembourg, **MC** Monaco, **NL** Netherlands, **PT** Portugal, **SE** Sweden, and any other State which is a Contracting State of the European Patent Convention and of the PCT
- ☑ **OA** **OAPI Patent:** **BF** Burkina Faso, **BJ** Benin, **CF** Central African Republic, **CG** Congo, **CI** Côte d'Ivoire, **CM** Cameroon, **GA** Gabon, **GN** Guinea, **ML** Mali, **MR** Mauritania, **NE** Niger, **SN** Senegal, **TD** Chad, **TG** Togo, and any other State which is a member State of OAPI and a Contracting State of the PCT *(if any other kind of protection or treatment desired, please specify on dotted line)*

National Patent *(if other kind of protection or treatment desired, specify on dotted line)*:

- ☑ **AL** Albania
- ☑ **AM** Armenia
- ☑ **AT** Austria
- ☑ **AU** Australia
- ☑ **AZ** Azerbaijan
- ☑ **BA** Bosnia and Herzegovina
- ☑ **BB** Barbados
- ☑ **BG** Bulgaria
- ☑ **BR** Brazil
- ☑ **BY** Belarus
- ☑ **CA** Canada
- ☑ **CH AND LI** Switzerland and Liechtenstein
- ☑ **CN** China
- ☑ **CU** Cuba
- ☑ **CZ** Czech Republic
- ☑ **DE** Germany
- ☑ **DK** Denmark
- ☑ **EE** Estonia
- ☑ **ES** Spain
- ☑ **FI** Finland
- ☑ **GB** United Kingdom
- ☑ **GE** Georgia
- ☑ **GH** Ghana
- ☑ **GM** Gambia
- ☑ **GW** Guinea-Bissau
- ☑ **HU** Hungary
- ☑ **ID** Indonesia
- ☑ **IL** Israel
- ☑ **IS** Iceland
- ☑ **JP** Japan
- ☑ **KE** Kenya
- ☑ **KG** Krygyzstan
- ☑ **KP** Democratic People's Republic of Korea
- ☑ **KR** Republic of Korea
- ☑ **KZ** Kazakstan
- ☑ **LC** Saint Lucia
- ☑ **LK** Sri Lanka
- ☑ **LR** Liberia
- ☑ **LS** Lesotho
- ☑ **LT** Lithuania
- ☑ **LU** Luxembourg
- ☑ **LV** Latvia
- ☑ **MD** Republic of Moldova
- ☑ **MG** Madagascar
- ☑ **MK** The former Yugoslav Republic of Macedonia
- ☑ **MN** Mongolia
- ☑ **MW** Malawi
- ☑ **MX** Mexico
- ☑ **NO** Norway
- ☑ **NZ** New Zealand
- ☑ **PL** Poland
- ☑ **PT** Portugal
- ☑ **RO** Romania
- ☑ **RU** Russian Federation
- ☑ **SD** Sudan
- ☑ **SE** Sweden
- ☑ **SG** Singapore
- ☑ **SI** Slovenia
- ☑ **SK** Slovakia
- ☑ **SL** Sierra Leone
- ☑ **TJ** Tajikstan
- ☑ **TM** Turkmenistan
- ☑ **TR** Turkey
- ☑ **TT** Trinidad and Tobago
- ☑ **UA** Ukraine
- ☑ **UG** Uganda
- ☑ **US** United States of America
- ☑ **UZ** Uzbekistan
- ☑ **VN** Viet Nam
- ☑ **YU** Yugoslavia
- ☑ **ZW** Zimbabwe

Check-boxes reserved for designating States (for the purposes of a national patent) which have become party to the PCT after the issuance of this sheet:

- ☑ **HR** Croatia
- ☑ **GD** Grenada
- ☑ **IN** India
- X **ZA** South Africa
- X **AE** United Arab Emirates

In addition to the designations made above, the applicant also makes under Rule 4.9(b) all designations which would be permitted under the PCT except the designation(s) of

The applicant declares that those additional designations are subject to confirmation and that any designation which is not confirmed before the expiration of 15 months from the priority date is to be regarded as withdrawn by the applicant at the expiration of that time limit.
(Confirmation of a designation consists of the filing of a notice specifying that designation and the payment of the designation and confirmation fees. Confirmation must reach the receiving Office within the 15-month time limit.)

Box No. VI PRIORITY CLAIM

Further priority claims are indicated in the Supplemental Box ☐

The priority of the following earlier application(s) is hereby claimed:

	Country *(in which, or for which, the application was filed)*	Filing Date *(day/month/year)*	Application No.	Office of filing *(only for regional or international application)*
item (1)	United Kingdom	17 Mar 1998	9805576.7	
item (2)	United Kingdom	31 Mar 1998	9806895.0	
item (3)	United Kingdom	3 Apr 1998	9807246.5	

Mark the following check-box if the certified copy of the earlier application is to be issued by the Office which for the purposes of the present international application is the receiving Office (a fee may be required):

☑ The receiving Office is hereby requested to prepare and transmit to the International Bureau a certified copy of the earlier application(s) identified above as item(s) : (1), (2), (3)

Box No. VII INTERNATIONAL SEARCHING AUTHORITY

Choice of International Searching Authority (ISA) *(If two or more International Searching Authorities are competent to carry out the international search, indicate the Authority chosen; the two-letter code may be used):* ISA / EPO

Earlier Search *Fill in where a search (international, international-type or other) by the International Search Authority has already been carried out or requested and the Authority is now requested to base the international search, to the extent possible, on the results of that earlier search. Identify such search or request either by reference to the relevant application (or the translation thereof) or by reference to the search request:*

Country (or regional Office): Date *(day/month/year)* Number:

Box No. VII CHECK LIST

This international application contains the following number of sheets:

1. request	:	5	sheets
2. description	:	42	sheets
3. claims	:	5	sheets
4. abstract	:	1	sheets
5. drawings	:	8	sheets
Total	:	61	sheets

This international application is accompanied by the item(s) marked below:

1. ☐ separate signed power of attorney
2. ☐ copy of general power of attorney
3. ☐ statement explaining lack of signature
4. ☐ priority documents(s) *identified in Box No. VI as item(s):*
5. ☐ fee calculation sheet
6. ☐ separate indications concerning deposited microorganisms
7. ☐ nucleotide and/or amino acid sequence listing (diskette)
8. ☐ other *(specify):*

Figure No. _____ of the drawings (if any) should accompany the abstract when it is published

Box No. IX SIGNATURE OF APPLICANT OR AGENT

Next to each signature, indicate the name of the person signing and the capacity in which the person signs (if such capacity is not obvious from reading the request)

A. M

MASCHIO, Antonio

For receiving Office use only

1. Date of actual receipt of the purported international application:
2. Drawings: ☐ received: ☐ not received:
3. Corrected date of actual receipt due to later but timely received papers or drawings completing the purported international application:
4. Date of timely receipt of the required corrections under PCT Article 11(2):
5. International Searching Authority specified by the applicant: ISA /
6. ☐ Transmittal of search copy delayed until search fee paid

For International Bureau use only

Date of receipt of the record copy by the International Bureau:

PATENT COOPERATION TREATY

From the
INTERNATIONAL PRELIMINARY EXAMINING AUTHORITY

To:

PTO/PCT Rec'd 18 SEP 2000

MASCHIO, Antonio
D YOUNG & CO
21 New Fetter Lane
London EC4A 1DA
GRANDE BRETAGNE

PCT

NOTIFICATION OF TRANSMITTAL OF THE INTERNATIONAL PRELIMINARY EXAMINATION REPORT

(PCT Rule 71.1)

Date of mailing *(day/month/year)*	11.07.2000
Applicant's or agent's file reference P004287WO ATM LSB	**IMPORTANT NOTIFICATION**

International application No.	International filing date *(day/month/year)*	Priority date *(day/month/year)*
PCT/GB99/00816	17/03/1999	17/03/1998

Applicant
GENDAQ LIMITED et al.

1. The applicant is hereby notified that this International Preliminary Examining Authority transmits herewith the international preliminary examination report and its annexes, if any, established on the international application.

2. A copy of the report and its annexes, if any, is being transmitted to the International Bureau for communication to all the elected Offices.

3. Where required by any of the elected Offices, the International Bureau will prepare an English translation of the report (but not of any annexes) and will transmit such translation to those Offices.

4. **REMINDER**

The applicant must enter the national phase before each elected Office by performing certain acts (filing translations and paying national fees) within 30 months from the priority date (or later in some Offices) (Article 39(1)) (see also the reminder sent by the International Bureau with Form PCT/IB/301).

Where a translation of the international application must be furnished to an elected Office, that translation must contain a translation of any annexes to the international preliminary examination report. It is the applicant's responsibility to prepare and furnish such translation directly to each elected Office concerned.

For further details on the applicable time limits and requirements of the elected Offices, see Volume II of the PCT Applicant's Guide.

Name and mailing address of the IPEA/	Authorized officer
European Patent Office D-80298 Munich Tel. +49 89 2399 - 0 Tx: 523656 epmu d Fax: +49 89 2399 - 4465	Christensen, J Tel.+49 89 2399-8052

Form PCT/IPEA/416 (July 1992)

THIS PAGE BLANK

PATENT COOPERATION TREATY

REC'D 12 JUL 2000
WIPO PCT

PCT

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

(PCT Article 36 and Rule 70)

Applicant's or agent's file reference	FOR FURTHER ACTION	See Notification of Transmittal of International Preliminary Examination Report (Form PCT/IPEA/416)
P004287WO ATM LSB		
International application No.	International filing date *(day/month/year)*	Priority date *(day/month/year)*
PCT/GB99/00816	17/03/1999	17/03/1998
International Patent Classification (IPC) or national classification and IPC C12N15/12		
Applicant GENDAQ LIMITED et al.		

1. This international preliminary examination report has been prepared by this International Preliminary Examining Authority and is transmitted to the applicant according to Article 36.

2. This REPORT consists of a total of 7 sheets, including this cover sheet.

   ☐ This report is also accompanied by ANNEXES, i.e. sheets of the description, claims and/or drawings which have been amended and are the basis for this report and/or sheets containing rectifications made before this Authority (see Rule 70.16 and Section 607 of the Administrative Instructions under the PCT).

   These annexes consist of a total of sheets.

3. This report contains indications relating to the following items:

   - I ☒ Basis of the report
   - II ☐ Priority
   - III ☐ Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
   - IV ☒ Lack of unity of invention
   - V ☒ Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations suporting such statement
   - VI ☐ Certain documents cited
   - VII ☐ Certain defects in the international application
   - VIII ☒ Certain observations on the international application

Date of submission of the demand	Date of completion of this report
06/10/1999	11.07.2000
Name and mailing address of the international preliminary examining authority: European Patent Office D-80298 Munich Tel. +49 89 2399 - 0 Tx: 523656 epmu d Fax: +49 89 2399 - 4465	Authorized officer Vix, O Telephone No. +49 89 2399 7326

Form PCT/IPEA/409 (cover sheet) (January 1994)

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

International application No. PCT/GB99/00816

I. Basis of the report

1. This report has been drawn on the basis of (*substitute sheets which have been furnished to the receiving Office in response to an invitation under Article 14 are referred to in this report as "originally filed" and are not annexed to the report since they do not contain amendments.*):

Description, pages:

1-42 as originally filed

Claims, No.:

1-27 as originally filed

Drawings, sheets:

1/8-8/8 as originally filed

2. The amendments have resulted in the cancellation of:

☐ the description, pages:
☐ the claims, Nos.:
☐ the drawings, sheets:

3. ☐ This report has been established as if (some of) the amendments had not been made, since they have been considered to go beyond the disclosure as filed (Rule 70.2(c)):

4. Additional observations, if necessary:

IV. Lack of unity of invention

1. In response to the invitation to restrict or pay additional fees the applicant has:

☐ restricted the claims.

☐ paid additional fees.

☐ paid additional fees under protest.

☐ neither restricted nor paid additional fees.

2. ☒ This Authority found that the requirement of unity of invention is not complied and chose, according to Rule 68.1, not to invite the applicant to restrict or pay additional fees.

3. This Authority considers that the requirement of unity of invention in accordance with Rules 13.1, 13.2 and 13.3 is

   ☐ complied with.

   ☒ not complied with for the following reasons:

   see separate sheet

4. Consequently, the following parts of the international application were the subject of international preliminary examination in establishing this report:

   ☒ all parts.

   ☐ the parts relating to claims Nos. .

V. Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

1. Statement

Novelty (N)	Yes:	Claims	2,4-27
	No:	Claims	1,3
Inventive step (IS)	Yes:	Claims	2,4-27
	No:	Claims	1,3
Industrial applicability (IA)	Yes:	Claims	1-27
	No:	Claims	

2. Citations and explanations

   see separate sheet

VIII. Certain observations on the international application

The following observations on the clarity of the claims, description, and drawings or on the question whether the claims are fully supported by the description, are made:

se s parate sh et

Re Item IV

Lack of unity of invention

The common inventive concept between all claims (except claim 3) is the polypeptide binding properties to a modified DNA triplet containing a 5-meC but not to the equivalent unmodified base.
Claim 3 is independant and refers to a method for preparing a "Zinc finger" polypeptide binding to a modified DNA triplet containing a 5-meC. There is no mention that such a polypeptide does not bind the unmodified identical DNA sequence, therefore widening the scope of the original claim 1 which contains this limitation. Thus, claim 3 leads to a unity problem because of a lack of common inventive concept (as required by Rule 13.1 PCT).

Correspondingly, present claims do not relate to one invention but to two separate ones, namely:

(i) Invention 1 (claims 1-2 and 4-27) : a Zinc finger polypeptide desgined to bind to a modified DNA triplet containing a 5-meC but NOT to the unmodified identical DNA sequence.
(ii) Invention 2 (claim 3) : a method to prepare a Zinc finger domain targeting a modified DNA triplet containing a 5-meCwherein however the binding tothe correspondng non-modified DNA triplet is not excluded.

Re Item V

Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

1. Reference is made to the following documents:
   D1: WO 98 53059 A (MEDICAL RES COUNCIL ;ISALAN MARK (GB); CHOO YEN (GB); KLUG AARON () 26 November 1998 (1998-11-26)
   D2: WO 98 53058 A (MEDICAL RES COUNCIL ;ISALAN MARK (GB); CHOO YEN (GB); KLUG AARON () 26 November 1998 (1998-11-26)
   D3: WO 98 53060 A (MEDICAL RES COUNCIL ;ISALAN MARK (GB); CHOO YEN (GB); KLUG AARON () 26 November 1998 (1998-11-26)

2. Novelty and Inventive step (Art. 33(2)(3) PCT)

The invention 1 discloses a Zinc finger polypeptide binding to a target DNA sequence containing a 5-meC at the central position, but not to an identical sequence with an unmodified central C base. A method for modifying a Zinc finger polypeptide in order to achieve the above mentioned differentiated DNA binding property is also part of the invention.

2.1. The claim 1 attempts to define the subject-matter in terms of a result to be achieved. The Zinc finger polypeptide technical features necessary for achieving the specific binding result have not been added. At this stage, any already known Zinc finger domains of the prior art could be tested to check whether or not it possesses inherently this specific binding capacity. Such a known polypeptide would take away the novelty of claim 1.

2.2. Claim 3 is independant and refers to a DNA binding polypeptide binding to a modified DNA triplet containing a 5-meC. There is no mention that such a polypeptide does not bind the unmodified identical DNA sequence, therefore widening the scope of the original claim 1. As claim 3 is maintained without modifications, there is a unity problem in the present application because of a lack of common inventive concept. Furthermore, there is a problem of novelty and inventive step as document D1 (or also D2-D3) explicitly discloses a method that would allow the person of skill to design a Zinc finger domain targeting *any* desired DNA sequence (D1, page 2, line 27-30). In the present application, the unexpected technical effect lies in the binding "differentiation" between modified and unmodified identical sequences which is absent of the subject matter of claim 3.

2.3 Remaining claims 2, 4-27 are dealing with zinc finger polypeptide which bond to a target DNA sequence. Said claims contain technical features that render them novel and inventive over the prior art.

Re Item VIII

Certain observations on the international application

1. Relating to Article 6 PCT:
   Claim 1 refers to a Zinc finger polypeptide which is supposed to bind to a target DNA sequence containing a modified base but not to an identical sequence with the unmodified base. Such a claim defines only a desideratum that clearly lacks the technical features that would allow the skilled person to realise said wish (typically a "result to be achieved"). There is no mention of the characteristics of such a Zinc finger polypeptide and the type of modification of the DNA sequence. The examples in the description are exclusively dealing with methylation of DNA sequences (as mentioned in claims 3-4, 7-8). The modification techniques focus on newly designed Zinc finger structural features that should be able to adapt methylation modification on "standard" bases (and not other type of chemical modification). A method for such a "custom" design is detailed in claim 3, and it gives clear indications how to modify a zinc finger domain to adapt a specific methylation modification in a DNA sequence. In opposite, claim 1 and the technical features of the claims dealing with very general "modified base" do not enable the person skilled in the art to perform the invention over the whole area of the claimed invention without undue burden (due to the necessity to screen a very large number of zinc finger polypeptides libraries in order to narrow down a structural element capable to perform the specific recognition).

2. The numbering discussed in claims 3-4 and 6 is not clearly identified using the convention adopted in the description (page 7) or in claim 10. Similarly, the definition of the symbol "++" in claim 4 (mentioned in page 4 of the description) has not been incorporated in the claim.

3. Claim 21 refers to a Zif 268 protein that is only described by references to other publications mentioned in brackets.
   As a general rule, claims shall not, except where absolutely necessary, rely, in respect of the technical features of the invention, on references to the description or drawings (Rule 6.2(b) PCT). In the present case an important technical feature is made by a reference to the description of scientific articles, which could render the scope of the claim broader or unclear. Zinc finger domains of the Zif 268 protein has no clear definition using a clear technical feature such as a protein sequence listing.

4. Claim 22 refers to the second Zinc finger selected from the protein Zif 268. The same observations as in part 3 discussed above applies for claim 22.

5. The wordings "randomisation" and "selection" in claim 26 are too vague and do not clearly delimit the scope of the claim. These steps need to be clearly defined in terms of strategy adopted for the randomisation process (a "partial" or "total" randomisation at every amino-acid position would conduct to very different protein domain) and type of selection that would allow "to improve the characteristics" of the DNA binding protein. Such "characteristics" (most likely the distinct binding capacity) of the protein are not stated explicitly.

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/00, C07K 14/00 | A2

(11) International Publication Number: **WO 99/47656**

(43) International Publication Date: 23 September 1999 (23.09.99)

(21) International Application Number: PCT/GB99/00816

(22) International Filing Date: 17 March 1999 (17.03.99)

(30) Priority Data:

9805576.7	17 March 1998 (17.03.98)	GB
9806895.0	31 March 1998 (31.03.98)	GB
9807246.5	3 April 1998 (03.04.98)	GB

(71) Applicant *(for all designated States except US)*: MEDICAL RESEARCH COUNCIL [GB/GB]; 20 Park Crescent, London W1N 4AL (GB).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: CHOO, Yen [GR/GB]; MRC Laboratory of Molecular Biology, Medical Research Council, Hills Road, Cambridge CB2 2QH (GB). ISALAN, Mark [GB/GB]; 24 Shottfield Avenue, East Sheen, London SW14 8EA (GB).

(74) Agents: MASCHIO, Antonio et al.; D. Young & Co., 21 New Fetter Lane, London EC4A 1DA (GB).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published

Without international search report and to be republished upon receipt of that report.

(54) Title: NUCLEIC ACID BINDING PROTEINS

(57) Abstract

The invention provides a method for producing a zinc finger polypeptide which binds to a target nucleic acid sequence containing a modified base but not to an identical sequence containing an equivalent unmodified base.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

Nucleic Acid Binding Proteins

The present invention relates to DNA binding proteins. In particular, the invention relates to a method for designing a protein which is capable of binding to a defined methylated DNA sequence but not to an equivalent unmethylated DNA sequence.

Protein-nucleic acid recognition is a commonplace phenomenon which is central to a large number of biomolecular control mechanisms which regulate the functioning of eukaryotic and prokaryotic cells. For instance, protein-DNA interactions form the basis of the regulation of gene expression and are thus one of the subjects most widely studied by molecular biologists.

A wealth of biochemical and structural information explains the details of protein-DNA recognition in numerous instances, to the extent that general principles of recognition have emerged. Many DNA-binding proteins contain independently folded domains for the recognition of DNA, and these domains in turn belong to a large number of structural families, such as the leucine zipper, the "helix-turn-helix" and zinc finger families.

Despite the great variety of structural domains, the specificity of the interactions observed to date between protein and DNA most often derives from the complementarity of the surfaces of a protein α-helix and the major groove of DNA [Klug, (1993) Gene 135:83-92]. In light of the recurring physical interaction of α-helix and major groove, the tantalising possibility arises that the contacts between particular amino acids and DNA bases could be described by a simple set of rules; in effect a stereochemical recognition code which relates protein primary structure to binding-site sequence preference.

It is clear, however, that no code will be found which can describe DNA recognition by all DNA-binding proteins. The structures of numerous complexes show significant differences in the way that the recognition α-helices of DNA-binding proteins from different structural families interact with the major groove of DNA, thus precluding similarities in patterns of recognition. The majority of known DNA-binding motifs are not particularly versatile, and any codes which might emerge would likely describe binding to a very few related DNA sequences.

Even within each family of DNA-binding proteins, moreover, it has hitherto appeared that the deciphering of a code would be elusive. Due to the complexity of the protein-DNA interaction, there does not appear to be a simple "alphabetic" equivalence between the primary structures of protein and nucleic acid which specifies a direct amino acid to base relationship.

International patent application WO 96/06166 addresses this issue and provides a "syllabic" code which explains protein-DNA interactions for zinc finger nucleic acid binding proteins. A syllabic code is a code which relies on more than one feature of the binding protein to specify binding to a particular base, the features being combinable in the forms of "syllables", or complex instructions, to define each specific contact.

Our copending UK patent applications, GB 9710805.4, 9710806.2, 9710807.0, 9710808.8, 9710809.6, 9710810.4, 9710811.2 and 9710812.0 describe improved techniques for designing zinc finger polypeptides capable of binding desired nucleic acid sequences. In combination with selection procedures, such as phage display, set forth for example in WO 96/06166, these techniques enable the production of zinc finger polypeptides capable of recognising practically any desired sequence.

Zinc finger domains studied and produced to date are capable of binding to recognition sequences composed by any of four nucleic acid bases: A, C, G or T (U in RNA). However, the DNA of many organisms includes also a fifth base, 5-methylcytosine (5-meC or, in nucleotide sequences herein, M). 5-meC arises from specific methylation of cytosine, and is used to mark the genome or to increase its information content. 5-methylcytosine is well known to affect protein-DNA interactions, for instance inhibiting cleavage of DNA by certain restriction enzymes. In vertebrates, cytosine is frequently methylated when directly preceding guanine, as in the dinucleotide CpG. This type of methylation generally down-regulates vertebrate gene expression, and can also prevent the binding of many eukaryotic transcription factors to DNA. Yet the zinc finger transcription factors tested to date, Sp1 and YY1, are not affected by CpG methylation of

their DNA binding sites, suggesting that zinc fingers are incapable of discriminating between cytosine and 5-meC.

Since methylated cytosine bases are involved with many regulatory interactions in gene expression, and particularly in eukaryotic, including human, gene expression, the production of zinc finger polypeptides which specifically target methylated cytosine bases would be highly desirable. Such polypeptides, in order to be useful, must be able to differentiate DNA sequences in which cytosine is methylated to 5-meC from identical non-methylated sequences.

Further nucleic acid base modifications are known in the art. For example, brominated nucleosides are known, such as Br-dU. Being photolabile, brominated nucleosides are useful in the determination of DNA-protein complex structure. Br-dU containing oligonucleotides are also useful as probes, since antibodies are available which recognise Br-dU. Moreover, in antisense oligonucleotide chemistry, the use of backbone modifications to improve oligonucleotide stability is well known; for example, phosphorothioate and 2'-O methylation are commonplace. Such backbone-modified nucleosides, and other nucleosides, may also be C-5 modified. For example, C-5 propyne derivatives and C-5 methylpyrimidine nucleosides are known and used in antisense nucleic acid chemistry.

Specific detection of modified nucleotides, and preferential binding of DNA-binding proteins thereto, is desirable. However, agents which are capable of reliably targeting a protein to a modified nucleic acid in a sequence-specific manner are not available in the art.

Summary of the Invention

We have now determined that modified nucleosides can be specifically recognised, over unmodified equivalents, by zinc finger polypeptides in a sequence-dependent manner. The invention accordingly provides a method for producing a zinc finger polypeptide

which binds to a target nucleic acid sequence containing a modified nucleic acid base, but not to an identical sequence containing the equivalent unmodified base.

In the present invention, a "modified" base is a nucleic acid base other than A, C, G or T as they occur in DNA in nature. Thus, the term modified includes methylated bases, such as 5-meC which occurs naturally in DNA, and base analogues, including naturally-occurring analogues such as U and artificial analogues such as I, backbone-modified bases and other artificial nucleosides.

In a first embodiment, the invention provides a method for preparing a DNA binding polypeptide of the Cys2-His2 zinc finger class capable of binding to a DNA triplet in target DNA sequence comprising 5-meC as the central residue in the target DNA triplet, wherein binding to the 5-meC residue by an α-helical zinc finger DNA binding motif of the polypeptide is achieved by placing an Ala residue at position +3 of the α-helix of the zinc finger.

All of the DNA-binding residue positions of zinc fingers, as referred to herein, are numbered from the first residue in the α-helix of the finger, ranging from +1 to +9. "-1" refers to the residue in the framework structure immediately preceding the α-helix in a Cys2-His2 zinc finger polypeptide. Residues referred to as "++" are residues present in an adjacent (C-terminal) finger. Where there is no C-terminal adjacent finger, "++" interactions do not operate.

Cys2-His2 zinc finger binding proteins, as is well known in the art, bind to target nucleic acid sequences via α-helical zinc metal atom co-ordinated binding motifs known as zinc fingers. Each zinc finger in a zinc finger nucleic acid binding protein is responsible for determining binding to a nucleic acid triplet in a nucleic acid binding sequence. Preferably, there are 2 or more zinc fingers, for example 2, 3, 4, 5 or 6 zinc fingers, in each binding protein. Advantageously, there are 3 zinc fingers in each zinc finger binding protein.

The method of the present invention allows the production of what are essentially artificial DNA binding proteins. In these proteins, artificial analogues of amino acids may be used, to impart the proteins with desired properties or for other reasons. Thus, the term "amino acid", particularly in the context where "any amino acid" is referred to, means any sort of natural or artificial amino acid or amino acid analogue that may be employed in protein construction according to methods known in the art. Moreover, any specific amino acid referred to herein may be replaced by a functional analogue thereof, particularly an artificial functional analogue. The nomenclature used herein therefore specifically comprises within its scope functional analogues of the defined amino acids.

The α-helix of a zinc finger binding protein aligns antiparallel to the nucleic acid strand, such that the primary nucleic acid sequence is arranged 3' to 5' in order to correspond with the N terminal to C-terminal sequence of the zinc finger. Since nucleic acid sequences are conventionally written 5' to 3', and amino acid sequences N-terminus to C-terminus, the result is that when a nucleic acid sequence and a zinc finger protein are aligned according to convention, the primary interaction of the zinc finger is with the - strand of the nucleic acid, since it is this strand which is aligned 3' to 5'. These conventions are followed in the nomenclature used herein. It should be noted, however, that in nature certain fingers, such as finger 4 of the protein GLI, bind to the + strand of nucleic acid: see Suzuki *et al.*, (1994) NAR 22:3397-3405 and Pavletich and Pabo, (1993) Science 261:1701-1707. The incorporation of such fingers into DNA binding molecules according to the invention is envisaged.

The invention provides a solution to a problem hitherto unaddressed in the art, by permitting the rational design of polypeptides which will bind DNA triplets containing a 5-meC residue, but not identical triplets containing a C residue.

The present invention may be integrated with the rules set forth for zinc finger polypeptide design in our copending UK patent applications listed above. In a preferred aspect, therefore, the invention provides a method for preparing a DNA binding polypeptide of the Cys2-His2 zinc finger class capable of binding to a DNA triplet in target DNA sequence comprising 5-meC, but not to an identical triplet comprising

unmethylated C, wherein binding to each base of the triplet by an α-helical zinc finger DNA binding motif in the polypeptide is determined as follows:

a) if the 5' base in the triplet is G, then position +6 in the α-helix is Arg and/or position ++2 is Asp;

b) if the 5' base in the triplet is A, then position +6 in the α-helix is Gln or Glu and ++2 is not Asp;

c) if the 5' base in the triplet is T, then position +6 in the α-helix is Ser or Thr and position ++2 is Asp; or position +6 is a hydrophobic amino acid other than Ala;

d) if the 5' base in the triplet is C, then position +6 in the α-helix may be any amino acid, provided that position ++2 in the α-helix is not Asp;

e) if the central base in the triplet is G, then position +3 in the α-helix is His;

f) if the central base in the triplet is A, then position +3 in the α-helix is Asn;

g) if the central base in the triplet is T, then position +3 in the α-helix is Ala, Ser, Ile, Leu, Thr or Val; provided that if it is Ala, then one of the residues at -1 or +6 is a small residue;

h) if the central base in the triplet is 5-meC, then position +3 in the α-helix is Ala, Ser, Ile, Leu, Thr or Val; provided that if it is Ala, then one of the residues at -1 or +6 is a small residue;

i) if the 3' base in the triplet is G, then position -1 in the α-helix is Arg;

j) if the 3' base in the triplet is A, then position -1 in the α-helix is Gln and position +2 is Ala;

k) if the 3' base in the triplet is T, then position -1 in the α-helix is Asn; or position -1 is Gln and position +2 is Ser;

l) if the 3' base in the triplet is C, then position -1 in the α-helix is Asp and Position +1 is Arg.

The foregoing represents a set of rules which permits the design of a zinc finger binding protein specific for any given DNA sequence incorporating 5-meC.

A zinc finger binding motif is a structure well known to those in the art and defined in, for example, Miller *et al.*, (1985) EMBO J. 4:1609-1614; Berg (1988) PNAS (USA)

85:99-102; Lee *et al.*, (1989) Science 245:635-637; see International patent applications WO 96/06166 and WO 96/32475, corresponding to USSN 08/422,107, incorporated herein by reference.

In general, a preferred zinc finger framework has the structure:

(A) X_{0-2} **C** X_{1-5} **C** X_{9-14} **H** X_{3-6} $^{H}/_{C}$

where X is any amino acid, and the numbers in subscript indicate the possible numbers of residues represented by X.

In a preferred aspect of the present invention, zinc finger nucleic acid binding motifs may be represented as motifs having the following primary structure:

(B) X^a **C** X_{2-4} **C** X_{2-3} F X^c X X X X L X X **H** X X X^b **H** - linker

-1 1 2 3 4 5 6 7 8 9

wherein X (including X^a, X^b and X^c) is any amino acid. X_{2-4} and X_{2-3} refer to the presence of 2 or 4, or 2 or 3, amino acids, respectively. The Cys and His residues, which together co-ordinate the zinc metal atom, are marked in bold text and are usually invariant, as is the Leu residue at position +4 in the α-helix.

Modifications to this representation may occur or be effected without necessarily abolishing zinc finger function, by insertion, mutation or deletion of amino acids. For example it is known that the second His residue may be replaced by Cys (Krizek *et al.*, (1991) J. Am. Chem. Soc. 113:4518-4523) and that Leu at +4 can in some circumstances be replaced with Arg. The Phe residue before X_c may be replaced by any aromatic other than Trp. Moreover, experiments have shown that departure from the preferred structure and residue assignments for the zinc finger are tolerated and may even prove beneficial in binding to certain nucleic acid sequences. Even taking this into account, however, the general structure involving an α-helix co-ordinated by a zinc atom which contacts four Cys or His residues, does not alter. As used herein, structures (A)

and (B) above are taken as an exemplary structure representing all zinc finger structures of the Cys2-His2 type.

Preferably, X^a is $^{F}/_{Y}$-X or P-$^{F}/_{Y}$-X. In this context, X is any amino acid. Preferably, in this context X is E, K, T or S. Less preferred but also envisaged are Q, V, A and P. The remaining amino acids remain possible.

Preferably, X_{2-4} consists of two amino acids rather than four. The first of these amino acids may be any amino acid, but S, E, K, T, P and R are preferred. Advantageously, it is P or R. The second of these amino acids is preferably E, although any amino acid may be used.

Preferably, X^b is T or I.

Preferably, X^c is S or T.

Preferably, X_{2-3} is G-K-A, G-K-C, G-K-S or G-K-G. However, departures from the preferred residues are possible, for example in the form of M-R-N or M-R.

Preferably, the linker is T-G-E-K or T-G-E-K-P.

As set out above, the major binding interactions occur with amino acids -1, +3 and +6. Amino acids +4 and +7 are largely invariant. The remaining amino acids may be essentially any amino acids. Preferably, position +9 is occupied by Arg or Lys. Advantageously, positions +1, +5 and +8 are not hydrophobic amino acids, that is to say are not Phe, Trp or Tyr. Preferably, position ++2 is any amino acid, and preferably serine, save where its nature is dictated by its role as a ++2 amino acid for an N-terminal zinc finger in the same nucleic acid binding molecule.

In a most preferred aspect, therefore, bringing together the above, the invention allows the definition of every residue in a zinc finger DNA binding motif which will bind specifically to a given DNA triplet incorporating a 5-meC residue as the central residue in

the triplet. Where targeting of a 5-meC containing sequence is desired, therefore, a suitable zinc finger can be constructed selecting a binding site such that 5-meC occurs at the centre of at least one base triplet thereof.

The code provided by the present invention is not entirely rigid; certain choices are provided. For example, positions +1, +5 and +8 may have any amino acid allocation, whilst other positions may have certain options: for example, the present rules provide that, for binding to a central T residue, any one of Ala, Ser or Val may be used at +3. In its broadest sense, therefore, the present invention provides a very large number of proteins which are capable of binding to every defined target DNA triplet incorporating 5-meC as the central residue and thereby any DNA binding site incorporating 5-meC.

Preferably, however, the number of possibilities may be significantly reduced. For example, the non-critical residues +1, +5 and +8 may be occupied by the residues Lys, Thr and Gln respectively as a default option. In the case of the other choices, for example, the first-given option may be employed as a default. Thus, the code according to the present invention allows the design of a single, defined polypeptide (a "default" polypeptide) which will bind to its target triplet.

In a further aspect of the present invention, there is provided a method for preparing a DNA binding protein of the Cys2-His2 zinc finger class capable of binding to a target DNA sequence incorporating 5-meC, comprising the steps of:

a) selecting a model zinc finger domain from the group consisting of naturally occurring zinc fingers and consensus zinc fingers; and

b) mutating at least one of positions -1, +3, +6 (and ++2) of the finger as required by a method according to the present invention.

In general, naturally occurring zinc fingers may be selected from those fingers for which the DNA binding specificity is known. For example, these may be the fingers for which a crystal structure has been resolved: namely Zif 268 (Elrod-Erickson *et al.*, (1996)

Structure 4:1171-1180), GLI (Pavletich and Pabo, (1993) Science 261:1701-1707), Tramtrack (Fairall *et al.*, (1993) Nature 366:483-487) and YY1 (Houbaviy *et al.*, (1996) PNAS (USA) 93:13577-13582).

The naturally occurring zinc finger 2 in Zif 268 makes an excellent starting point from which to engineer a zinc finger and is preferred.

Consensus zinc finger structures may be prepared by comparing the sequences of known zinc fingers, irrespective of whether their binding domain is known. Preferably, the consensus structure is selected from the group consisting of the consensus structure P Y K C P E C G K S F S Q K S D L V K H Q R T H T G, and the consensus structure P Y K C S E C G K A F S Q K S N L T R H Q R I H T G E K P.

The consensuses are derived from the consensus provided by Krizek *et al.*, (1991) J. Am. Chem. Soc. 113:4518-4523 and from Jacobs, (1993) PhD thesis, University of Cambridge, UK. In both cases, the linker sequences described above for joining two zinc finger motifs together, namely TGEK or TGEKP can be formed on the ends of the consensus. Thus, a P may be removed where necessary, or, in the case of the consensus terminating T G, E K (P) can be added.

When the nucleic acid specificity of the model finger selected is known, the mutation of the finger in order to modify its specificity to bind to the target DNA may be directed to residues known to affect binding to bases at which the natural and desired targets differ. Otherwise, mutation of the model fingers should be concentrated upon residues -1, +3, +6 and ++2 as provided for in the foregoing rules.

In order to produce a binding protein having improved binding, moreover, the rules provided by the present invention may be supplemented by physical or virtual modelling of the protein/DNA interface in order to assist in residue selection.

In a second embodiment, the invention provides a method for producing a zinc finger polypeptide capable of binding to a DNA sequence comprising a modified residue, but not to an identical sequence comprising an equivalent unmodified residue, comprising:

a) providing a nucleic acid library encoding a repertoire of zinc finger polypeptides, the nucleic acid members of the library being at least partially randomised at one or more of the positions encoding residues -1, 2, 3 and 6 of the α-helix of the zinc finger polypeptides;

b) displaying the library in a selection system and screening it against a target DNA sequence comprising the modified residue;

c) isolating the nucleic acid members of the library encoding zinc finger polypeptides capable of binding to the target sequence; and

d) optionally, verifying that the zinc finger polypeptides do not bind significantly to a DNA sequence identical to the target DNA sequence but containing the unmodified residue in place of the modified residue.

Methods for the production of libraries encoding randomised polypeptides are known in the art and may be applied in the present invention. Randomisation may be total, or partial; in the case of partial randomisation, the selected codons preferably encode options for amino acids as set forth in the rules of the first embodiment of the present invention. Thus, the first and second embodiments may advantageously be combined.

Preferably, the modified residue is 5-meC and the unmodified residue is C. However, other modifications may be targeted by the method of the invention. For example, zinc finger polypeptides may be designed which specifically bind to nucleic acids incorporating the base U, in preference to the equivalent base T. An advantage of the second embodiment of the invention is that zinc finger polypeptides may be developed to bind to any DNA sequence incorporating a modified base, irrespective of its positioning in the target DNA triplet.

In a further preferred aspect, the invention comprises a method for producing a zinc finger polypeptide capable of binding to a DNA sequence comprising a modified residue, but not to an identical sequence comprising an equivalent unmodified residue, comprising:

a) providing a nucleic acid library encoding a repertoire of zinc finger polypeptides each possessing more than one zinc fingers, the nucleic acid members of the library being at least partially randomised at one or more of the positions encoding residues -1, 2, 3 and 6 of the α-helix in a first zinc finger and at one or more of the positions encoding residues -1, 2, 3 and 6 of the α-helix in a further zinc finger of the zinc finger polypeptides;

b) displaying the library in a selection system and screening it against a target DNA sequence comprising the modified residue;

c) isolating the nucleic acid members of the library encoding zinc finger polypeptides capable of binding to the target sequence; and

d) optionally, verifying that the zinc finger polypeptides do not bind significantly to a DNA sequence identical to the target DNA sequence but containing the unmodified residue in place of the modified residue.

In this aspect, the invention encompasses library technology described in our copending International patent application WO98/53057, incorporated herein by reference in its entirety. WO98/53057 describes the production of zinc finger polypeptide libraries in which each individual zinc finger polypeptide comprises more than one, for example two or three, zinc fingers; and wherein within each polypeptide partial randomisation occurs in at least two zinc fingers.

This allows for the selection of the "overlap" specificity, wherein, within each triplet, the choice of residue for binding to the third nucleotide (read 3' to 5' on the + strand) is influenced by the residue present at position +2 on the subsequent zinc finger, which

displays cross-strand specificity in binding. The selection of zinc finger polypeptides incorporating cross-strand specificity of adjacent zinc fingers enables the selection of nucleic acid binding proteins with a higher degree of specificity than is otherwise possible.

Advantageously, in order to derive the greatest benefit, the binding site is selected such that the modified base is in position 3 of one of the triplets, such that cross-strand specificity can be relied upon to contact the parallel strand in the corresponding position and introduce a further level of discrimination.

In a third embodiment, the present invention may be applied to the production of zinc finger polypeptides capable of binding to a DNA sequence comprising an unmethylated C residue, but not to an identical sequence comprising a 5-meC residue. This may be carried out by differential screening, as set forth above. Moreover, rules may be applied in addition to or instead of screening.

Where the central residue of a target triplet is C, the use of Asp at position +3 of a zinc finger polypeptide allows preferential binding to C over 5-meC.

Brief Description of the Figures

Figure 1a is an alignment of the amino acid sequence of the three fingers from Zif268 used in a phage display library. Randomised residue positions in the α-helix of finger 2 are marked 'X' and are numbered above the alignment relative to the first helical residue (position +1). Residues which form the hydrophobic core are circled; zinc ligands are written as white letters on a black circle background; and positions comprising the secondary structure elements of a zinc finger are marked below the sequence.

Figure 1b shows amino acid sequences of the variant α-helical regions from some zinc fingers selected by phage display using the DNA binding site GCGGNGGCG where the central (bold) nucleotide of the middle (underlined) triplet was either: (i) 5-

methylcytosine, (ii) thymine, or (iii) cytosine. Amino acid sequences are listed below the DNA oligonucleotide used in their selection. Amino acid positions are numbered above the aligned sequences relative to the first helical residue (position +1). Circled residues (in position +3) are predicted to contact the middle nucleotide of the binding site.

Figure 1c shows a phage ELISA binding assay showing discrimination of pyrimidines by representative phage-selected zinc fingers. The matrix shows three different zinc finger phage clones (x, y and z) reacted with four different DNA binding sites present at a concentration of 3nM. Binding is represented by vertical bars which indicate the OD obtained by ELISA (Choo and Klug, (1997) Curr. Opin. Str. Biol. 7:117-125). The amino acid sequences of the variant α-helical regions from the selected zinc fingers are: REDVLIRHGK (x), RADALMVHKR (y), and RGPDLARHGR (z). The DNA sequences contain the generic binding site GCGGNGGCG, where the central (bold) nucleotide was either: uracil (U), thymine (T), cytosine (C), or 5-methylcytosine (M).

Figure 2 shows the effect of cytosine methylation on DNA binding by phage-selected zinc fingers. Graphs show three different zinc finger phage binding to the DNA sequence GCGGCGGCG in the presence (circle) and absence (triangle) of methylation of the central base (bold). The zinc finger clones tested contained variant α-helical regions of the middle finger as follows: (a) RADALMVHKR, (b) RGPDLARHGR and (c) REDVLIRHGK. These respective zinc finger clones preferentially bind their cognate DNA site in the presence, absence, or regardless of cytosine methylation.

Figure 3 shows the binding site interactions of 5 zinc finger polypeptides, selected taking into account cross-strand specificity by overlapping finger randomisation, with each of the oligonucleotides used in the selection process. Cross-strand contacts are shown.

Figure 4 is analogous to Figure 2 and shows the binding curves for four of the polypeptides as described in Figure 3 to their respective oligonucleotides.

Figure 5 shows discrimination between 5-meC and T by zfHAE(M).

Figure 6 shows binding of zinc finger polypeptides zfHHA(M) and zfHAE(M) to a nucleotide sequence (Figure 6a) in response to selective methylation by addition of methylase enzymes (Figure 6b). Polypeptides zfHHA(Y) and zfHAE(Y) do not discriminate between methylated and unmethylated DNA, as expected.

Detailed Description of the Invention

Randomisation involves may involve of zinc finger polypeptides at the DNA or protein level. Mutagenesis and screening of zinc finger polypeptides may be achieved by any suitable means. Preferably, the mutagenesis is performed at the nucleic acid level, for example by synthesising novel genes encoding mutant proteins and expressing these to obtain a variety of different proteins. Alternatively, existing genes can be themselves mutated, such by site-directed or random mutagenesis, in order to obtain the desired mutant genes.

Mutations may be performed by any method known to those of skill in the art. Preferred, however, is site-directed mutagenesis of a nucleic acid sequence encoding the protein of interest. A number of methods for site-directed mutagenesis are known in the art, from methods employing single-stranded phage such as M13 to PCR-based techniques (see "PCR Protocols: A guide to methods and applications", M.A. Innis, D.H. Gelfand, J.J. Sninsky, T.J. White (eds.). Academic Press, New York, 1990). Preferably, the commercially available Altered Site II Mutagenesis System (Promega) may be employed, according to the directions given by the manufacturer.

Randomisation of the zinc finger binding motifs produced according to the invention is preferably directed to those residues where the code provided herein gives a choice of residues. For example, therefore, positions +1, +5 and +8 are advantageously randomised, whilst preferably avoiding hydrophobic amino acids; positions involved in binding to the nucleic acid, notably -1, +2, +3 and +6, may be randomised also, preferably within the choices provided by the rules of the present invention.

Screening of the proteins produced by mutant genes is preferably performed by expressing the genes and assaying the binding ability of the protein product. A simple and advantageously rapid method by which this may be accomplished is by phage display, in which the mutant polypeptides are expressed as fusion proteins with the coat proteins of filamentous bacteriophage, such as the minor coat protein pII of bacteriophage m13 or gene III of bacteriophage Fd, and displayed on the capsid of bacteriophage transformed with the mutant genes. The target nucleic acid sequence is used as a probe to bind directly to the protein on the phage surface and select the phage possessing advantageous mutants, by affinity purification. The phage are then amplified by passage through a bacterial host, and subjected to further rounds of selection and amplification in order to enrich the mutant pool for the desired phage and eventually isolate the preferred clone(s). Detailed methodology for phage display is known in the art and set forth, for example, in US Patent 5,223,409; Choo and Klug, (1995) Current Opinions in Biotechnology 6:431-436; Smith, (1985) Science 228:1315-1317; and McCafferty *et al.*, (1990) Nature 348:552-554; all incorporated herein by reference. Vector systems and kits for phage display are available commercially, for example from Pharmacia.

Specific peptide ligands such as zinc finger polypeptides may moreover be selected for binding to targets by affinity selection using large libraries of peptides linked to the C terminus of the lac repressor LacI (Cull *et al.*, (1992) Proc Natl Acad Sci U S A, 89, 1865-9). When expressed in *E. coli* the repressor protein physically links the ligand to the encoding plasmid by binding to a lac operator sequence on the plasmid.

An entirely *in vitro* polysome display system has also been reported (Mattheakis *et al.*, (1994) Proc Natl Acad Sci U S A, 91, 9022-6) in which nascent peptides are physically attached via the ribosome to the RNA which encodes them.

The library of the invention may randomised at those positions for which choices are given in the rules of the first embodiment of the present invention. In particular, the members of the library are randomised at position +3 for binding to a central 5-meC residue. In such a case, 5-meC binding polypeptides will be selected by comparative binding analyses against methylated and non-methylated binding sites. However, the

rules set forth above allow the person of ordinary skill in the art to make informed choices concerning the desired codon usage at the given positions. For instance, position +3 in the case of a central 5-meC residue should be Ala residue, encoded by the codon GCN.

Zinc finger binding motifs designed according to the invention may be combined into nucleic acid binding proteins having a multiplicity of zinc fingers. Preferably, the proteins have at least two zinc fingers. In nature, zinc finger binding proteins commonly have at least three zinc fingers, although two-zinc finger proteins such as Tramtrack are known. The presence of at least three zinc fingers is preferred. Binding proteins may be constructed by joining the required fingers end to end, N-terminus to C-terminus. Preferably, this is effected by joining together the relevant nucleic acid coding sequences encoding the zinc fingers to produce a composite coding sequence encoding the entire binding protein. The invention therefore provides a method for producing a DNA binding protein as defined above, wherein the DNA binding protein is constructed by recombinant DNA technology, the method comprising the steps of:

a) preparing a nucleic acid coding sequence encoding two or more zinc finger binding motifs as defined above, placed N-terminus to C-terminus;
b) inserting the nucleic acid sequence into a suitable expression vector; and
c) expressing the nucleic acid sequence in a host organism in order to obtain the DNA binding protein.

A "leader" peptide may be added to the N-terminal finger. Preferably, the leader peptide is MAEEKP.

The nucleic acid encoding the DNA binding protein according to the invention can be incorporated into vectors for further manipulation. As used herein, vector (or plasmid) refers to discrete elements that are used to introduce heterologous nucleic acid into cells for either expression or replication thereof. Selection and use of such vehicles are well within the skill of the person of ordinary skill in the art. Many vectors are available, and selection of appropriate vector will depend on the intended use of the vector, i.e. whether

it is to be used for DNA amplification or for nucleic acid expression, the size of the DNA to be inserted into the vector, and the host cell to be transformed with the vector. Each vector contains various components depending on its function (amplification of DNA or expression of DNA) and the host cell for which it is compatible. The vector components generally include, but are not limited to, one or more of the following: an origin of replication, one or more marker genes, an enhancer element, a promoter, a transcription termination sequence and a signal sequence.

Both expression and cloning vectors generally contain nucleic acid sequence that enable the vector to replicate in one or more selected host cells. Typically in cloning vectors, this sequence is one that enables the vector to replicate independently of the host chromosomal DNA, and includes origins of replication or autonomously replicating sequences. Such sequences are well known for a variety of bacteria, yeast and viruses. The origin of replication from the plasmid pBR322 is suitable for most Gram-negative bacteria, the 2μ plasmid origin is suitable for yeast, and various viral origins (e.g. SV 40, polyoma, adenovirus) are useful for cloning vectors in mammalian cells. Generally, the origin of replication component is not needed for mammalian expression vectors unless these are used in mammalian cells competent for high level DNA replication, such as COS cells.

Most expression vectors are shuttle vectors, i.e. they are capable of replication in at least one class of organisms but can be transfected into another class of organisms for expression. For example, a vector is cloned in *E. coli* and then the same vector is transfected into yeast or mammalian cells even though it is not capable of replicating independently of the host cell chromosome. DNA may also be replicated by insertion into the host genome. However, the recovery of genomic DNA encoding the DNA binding protein is more complex than that of exogenously replicated vector because restriction enzyme digestion is required to excise DNA binding protein DNA. DNA can be amplified by PCR and be directly transfected into the host cells without any replication component.

Advantageously, an expression and cloning vector may contain a selection gene also referred to as selectable marker. This gene encodes a protein necessary for the survival or growth of transformed host cells grown in a selective culture medium. Host cells not transformed with the vector containing the selection gene will not survive in the culture medium. Typical selection genes encode proteins that confer resistance to antibiotics and other toxins, e.g. ampicillin, neomycin, methotrexate or tetracycline, complement auxotrophic deficiencies, or supply critical nutrients not available from complex media.

As to a selective gene marker appropriate for yeast, any marker gene can be used which facilitates the selection for transformants due to the phenotypic expression of the marker gene. Suitable markers for yeast are, for example, those conferring resistance to antibiotics G418, hygromycin or bleomycin, or provide for prototrophy in an auxotrophic yeast mutant, for example the URA3, LEU2, LYS2, TRP1, or HIS3 gene.

Since the replication of vectors is conveniently done in *E. coli*, an *E. coli* genetic marker and an *E. coli* origin of replication are advantageously included. These can be obtained from *E. coli* plasmids, such as pBR322, Bluescript© vector or a pUC plasmid, e.g. pUC18 or pUC19, which contain both *E. coli* replication origin and *E. coli* genetic marker conferring resistance to antibiotics, such as ampicillin.

Suitable selectable markers for mammalian cells are those that enable the identification of cells competent to take up DNA binding protein nucleic acid, such as dihydrofolate reductase (DHFR, methotrexate resistance), thymidine kinase, or genes conferring resistance to G418 or hygromycin. The mammalian cell transformants are placed under selection pressure which only those transformants which have taken up and are expressing the marker are uniquely adapted to survive. In the case of a DHFR or glutamine synthase (GS) marker, selection pressure can be imposed by culturing the transformants under conditions in which the pressure is progressively increased, thereby leading to amplification (at its chromosomal integration site) of both the selection gene and the linked DNA that encodes the DNA binding protein. Amplification is the process by which genes in greater demand for the production of a protein critical for growth, together with closely associated genes which may encode a desired protein, are reiterated

in tandem within the chromosomes of recombinant cells. Increased quantities of desired protein are usually synthesised from thus amplified DNA.

Expression and cloning vectors usually contain a promoter that is recognised by the host organism and is operably linked to DNA binding protein encoding nucleic acid. Such a promoter may be inducible or constitutive. The promoters are operably linked to DNA encoding the DNA binding protein by removing the promoter from the source DNA by restriction enzyme digestion and inserting the isolated promoter sequence into the vector. Both the native DNA binding protein promoter sequence and many heterologous promoters may be used to direct amplification and/or expression of DNA binding protein encoding DNA.

Promoters suitable for use with prokaryotic hosts include, for example, the β-lactamase and lactose promoter systems, alkaline phosphatase, the tryptophan (trp) promoter system and hybrid promoters such as the tac promoter. Their nucleotide sequences have been published, thereby enabling the skilled worker operably to ligate them to DNA encoding DNA binding protein, using linkers or adapters to supply any required restriction sites. Promoters for use in bacterial systems will also generally contain a Shine-Delgarno sequence operably linked to the DNA encoding the DNA binding protein.

Preferred expression vectors are bacterial expression vectors which comprise a promoter of a bacteriophage such as phagex or T7 which is capable of functioning in the bacteria. In one of the most widely used expression systems, the nucleic acid encoding the fusion protein may be transcribed from the vector by T7 RNA polymerase (Studier et al, Methods in Enzymol. 185; 60-89, 1990). In the *E. coli* BL21(DE3) host strain, used in conjunction with pET vectors, the T7 RNA polymerase is produced from the λ-lysogen DE3 in the host bacterium, and its expression is under the control of the IPTG inducible lac UV5 promoter. This system has been employed successfully for over-production of many proteins. Alternatively the polymerase gene may be introduced on a lambda phage by infection with an int- phage such as the CE6 phage which is commercially available (Novagen, Madison, USA). other vectors include vectors containing the lambda PL promoter such as PLEX (Invitrogen, NL) , vectors containing the trc promoters such as

pTrcHisXpressTm (Invitrogen) or pTrc99 (Pharmacia Biotech, SE) or vectors containing the tac promoter such as pKK223-3 (Pharmacia Biotech) or PMAL (New England Biolabs, MA, USA).

Moreover, the DNA binding protein gene according to the invention preferably includes a secretion sequence in order to facilitate secretion of the polypeptide from bacterial hosts, such that it will be produced as a soluble native peptide rather than in an inclusion body. The peptide may be recovered from the bacterial periplasmic space, or the culture medium, as appropriate.

Suitable promoting sequences for use with yeast hosts may be regulated or constitutive and are preferably derived from a highly expressed yeast gene, especially a Saccharomyces cerevisiae gene. Thus, the promoter of the TRP1 gene, the ADHI or ADHII gene, the acid phosphatase (PH05) gene, a promoter of the yeast mating pheromone genes coding for the a- or α-factor or a promoter derived from a gene encoding a glycolytic enzyme such as the promoter of the enolase, glyceraldehyde-3-phosphate dehydrogenase (GAP), 3-phospho glycerate kinase (PGK), hexokinase, pyruvate decarboxylase, phosphofructokinase, glucose-6-phosphate isomerase, 3-phosphoglycerate mutase, pyruvate kinase, triose phosphate isomerase, phosphoglucose isomerase or glucokinase genes, or a promoter from the TATA binding protein (TBP) gene can be used. Furthermore, it is possible to use hybrid promoters comprising upstream activation sequences (UAS) of one yeast gene and downstream promoter elements including a functional TATA box of another yeast gene, for example a hybrid promoter including the UAS(s) of the yeast PH05 gene and downstream promoter elements including a functional TATA box of the yeast GAP gene (PH05-GAP hybrid promoter). A suitable constitutive PHO5 promoter is e.g. a shortened acid phosphatase PH05 promoter devoid of the upstream regulatory elements (UAS) such as the PH05 (-173) promoter element starting at nucleotide -173 and ending at nucleotide -9 of the PH05 gene.

DNA binding protein gene transcription from vectors in mammalian hosts may be controlled by promoters derived from the genomes of viruses such as polyoma virus,

adenovirus, fowlpox virus, bovine papilloma virus, avian sarcoma virus, cytomegalovirus (CMV), a retrovirus and Simian Virus 40 (SV40), from heterologous mammalian promoters such as the actin promoter or a very strong promoter, e.g. a ribosomal protein promoter, and from the promoter normally associated with DNA binding protein sequence, provided such promoters are compatible with the host cell systems.

Transcription of a DNA encoding DNA binding protein by higher eukaryotes may be increased by inserting an enhancer sequence into the vector. Enhancers are relatively orientation and position independent. Many enhancer sequences are known from mammalian genes (e.g. elastase and globin). However, typically one will employ an enhancer from a eukaryotic cell virus. Examples include the SV40 enhancer on the late side of the replication origin (bp 100-270) and the CMV early promoter enhancer. The enhancer may be spliced into the vector at a position 5' or 3' to DNA binding protein DNA, but is preferably located at a site 5' from the promoter.

Advantageously, a eukaryotic expression vector encoding a DNA binding protein according to the invention may comprise a locus control region (LCR). LCRs are capable of directing high-level integration site independent expression of transgenes integrated into host cell chromatin, which is of importance especially where the DNA binding protein gene is to be expressed in the context of a permanently-transfected eukaryotic cell line in which chromosomal integration of the vector has occurred, or in transgenic animals.

Eukaryotic vectors may also contain sequences necessary for the termination of transcription and for stabilising the mRNA. Such sequences are commonly available from the 5' and 3' untranslated regions of eukaryotic or viral DNAs or cDNAs. These regions contain nucleotide segments transcribed as polyadenylated fragments in the untranslated portion of the mRNA encoding DNA binding protein.

An expression vector includes any vector capable of expressing DNA binding protein nucleic acids that are operatively linked with regulatory sequences, such as promoter regions, that are capable of expression of such DNAs. Thus, an expression vector refers

to a recombinant DNA or RNA construct, such as a plasmid, a phage, recombinant virus or other vector, that upon introduction into an appropriate host cell, results in expression of the cloned DNA. Appropriate expression vectors are well known to those with ordinary skill in the art and include those that are replicable in eukaryotic and/or prokaryotic cells and those that remain episomal or those which integrate into the host cell genome. For example, DNAs encoding DNA binding protein may be inserted into a vector suitable for expression of cDNAs in mammalian cells, e.g. a CMV enhancer-based vector such as pEVRF (Matthias, et al., (1989) NAR 17, 6418).

Particularly useful for practising the present invention are expression vectors that provide for the transient expression of DNA encoding DNA binding protein in mammalian cells. Transient expression usually involves the use of an expression vector that is able to replicate efficiently in a host cell, such that the host cell accumulates many copies of the expression vector, and, in turn, synthesises high levels of DNA binding protein. For the purposes of the present invention, transient expression systems are useful e.g. for identifying DNA binding protein mutants, to identify potential phosphorylation sites, or to characterise functional domains of the protein.

Construction of vectors according to the invention employs conventional ligation techniques. Isolated plasmids or DNA fragments are cleaved, tailored, and religated in the form desired to generate the plasmids required. If desired, analysis to confirm correct sequences in the constructed plasmids is performed in a known fashion. Suitable methods for constructing expression vectors, preparing in vitro transcripts, introducing DNA into host cells, and performing analyses for assessing DNA binding protein expression and function are known to those skilled in the art. Gene presence, amplification and/or expression may be measured in a sample directly, for example, by conventional Southern blotting, Northern blotting to quantitate the transcription of mRNA, dot blotting (DNA or RNA analysis), or in situ hybridisation, using an appropriately labelled probe which may be based on a sequence provided herein. Those skilled in the art will readily envisage how these methods may be modified, if desired.

In accordance with another embodiment of the present invention, there are provided cells containing the above-described nucleic acids. Such host cells such as prokaryote, yeast and higher eukaryote cells may be used for replicating DNA and producing the DNA binding protein. Suitable prokaryotes include eubacteria, such as Gram-negative or Gram-positive organisms, such as *E. coli*, e.g. *E. coli* K-12 strains, DH5α and HB101, or Bacilli. Further hosts suitable for the DNA binding protein encoding vectors include eukaryotic microbes such as filamentous fungi or yeast, e.g. Saccharomyces cerevisiae. Higher eukaryotic cells include insect and vertebrate cells, particularly mammalian cells including human cells, or nucleated cells from other multicellular organisms. In recent years propagation of vertebrate cells in culture (tissue culture) has become a routine procedure. Examples of useful mammalian host cell lines are epithelial or fibroblastic cell lines such as Chinese hamster ovary (CHO) cells, NIH 3T3 cells, HeLa cells or 293T cells. The host cells referred to in this disclosure comprise cells in *in vitro* culture as well as cells that are within a host animal.

DNA may be stably incorporated into cells or may be transiently expressed using methods known in the art. Stably transfected mammalian cells may be prepared by transfecting cells with an expression vector having a selectable marker gene, and growing the transfected cells under conditions selective for cells expressing the marker gene. To prepare transient transfectants, mammalian cells are transfected with a reporter gene to monitor transfection efficiency.

To produce such stably or transiently transfected cells, the cells should be transfected with a sufficient amount of the DNA binding protein-encoding nucleic acid to form the DNA binding protein. The precise amounts of DNA encoding the DNA binding protein may be empirically determined and optimised for a particular cell and assay.

Host cells are transfected or, preferably, transformed with the above-captioned expression or cloning vectors of this invention and cultured in conventional nutrient media modified as appropriate for inducing promoters, selecting transformants, or amplifying the genes encoding the desired sequences. Heterologous DNA may be introduced into host cells by any method known in the art, such as transfection with a vector encoding a heterologous

DNA by the calcium phosphate coprecipitation technique or by electroporation. Numerous methods of transfection are known to the skilled worker in the field. Successful transfection is generally recognised when any indication of the operation of this vector occurs in the host cell. Transformation is achieved using standard techniques appropriate to the particular host cells used.

Incorporation of cloned DNA into a suitable expression vector, transfection of eukaryotic cells with a plasmid vector or a combination of plasmid vectors, each encoding one or more distinct genes or with linear DNA, and selection of transfected cells are well known in the art (see, e.g. Sambrook et al. (1989) Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor Laboratory Press).

Transfected or transformed cells are cultured using media and culturing methods known in the art, preferably under conditions, whereby the DNA binding protein encoded by the DNA is expressed. The composition of suitable media is known to those in the art, so that they can be readily prepared. Suitable culturing media are also commercially available.

DNA binding proteins according to the invention may be employed in a wide variety of applications, including diagnostics and as research tools. Advantageously, they may be employed as diagnostic tools for identifying the presence of modified nucleic acid molecules in a complex mixture. DNA binding molecules according to the invention can differentiate single base modifications in target DNA molecules.

For example, zinc fingers may be fused to nucleic acid cleavage moieties, such as the catalytic domain of a restriction enzyme, to produce a restriction enzyme capable of cleaving only methylated DNA (see Kim, *et al.*, (1996) Proc. Natl. Acad. Sci. USA 93:1156-1160) . Using such approaches, different zinc finger domains can be used to create restriction enzymes with any desired recognition nucleotide sequence, but which cleave DNA conditionally dependent on the particular modification of the nucleotides, for instance methylation of the cytosine ring at position 5.

5-meC targeting zinc fingers may moreover be employed in the regulation of gene transcription, for example by specific cleavage of methylated (or unmethylated) sequences using a fusion polypeptide comprising a zinc finger targeting domain and a DNA cleavage domain, or by fusion of an activating domain (such as HSV VP16) to a zinc finger, to activate transcription from a gene which possesses the zinc finger binding sequence in its upstream sequences. Activation only occurs when the target DNA is modified, such as by methylation. Zinc fingers capable of differentiating between U and T may be used to preferentially target RNA or DNA, as required. Where RNA-targeting polypeptides are intended, these are included in the term "DNA-binding molecule".

In a preferred embodiment, the zinc finger polypeptides of the invention may be employed to detect the presence of a particular base modification in a target nucleic acid sequence in a sample.

Accordingly, the invention provides a method for determining the presence of a target modified nucleic acid molecule, comprising the steps of:

a) preparing a DNA binding protein by the method set forth above which is specific for the target modified nucleic acid molecule;
b) exposing a test system comprising the target modified nucleic acid molecule to the DNA binding protein under conditions which promote binding, and removing any DNA binding protein which remains unbound;
c) detecting the presence of the DNA binding protein in the test system.

In a preferred embodiment, the DNA binding molecules of the invention can be incorporated into an ELISA assay. For example, phage displaying the molecules of the invention can be used to detect the presence of the target DNA, and visualised using enzyme-linked anti-phage antibodies.

Further improvements to the use of zinc finger phage for diagnosis can be made, for example, by co-expressing a marker protein fused to the minor coat protein (gVIII) of bacteriophage. Since detection with an anti-phage antibody would then be obsolete, the

time and cost of each diagnosis would be further reduced. Depending on the requirements, suitable markers for display might include the fluorescent proteins (A. B. Cubitt, *et al.*, (1995) *Trends Biochem Sci.* **20**, 448-455; T. T. Yang, *et al.*, (1996) *Gene* **173**, 19-23), or an enzyme such as alkaline phosphatase which has been previously displayed on gIII (J. McCafferty, R. H. Jackson, D. J. Chiswell, (1991) *Protein Engineering* **4**, 955-961) Labelling different types of diagnostic phage with distinct markers would allow multiplex screening of a single DNA sample. Nevertheless, even in the absence of such refinements, the basic ELISA technique is reliable, fast, simple and particularly inexpensive. Moreover it requires no specialised apparatus, nor does it employ hazardous reagents such as radioactive isotopes, making it amenable to routine use in the clinic. The major advantage of the protocol is that it obviates the requirement for gel electrophoresis, and so opens the way to automated DNA diagnosis.

The invention provides DNA binding proteins which can be engineered with exquisite specificity. The invention lends itself, therefore, to the design of any molecule of which specific DNA binding is required. For example, the proteins according to the invention may be employed in the manufacture of chimeric restriction enzymes, in which a nucleic acid cleaving domain is fused to a DNA binding domain comprising a zinc finger as described herein.

The invention is described below, for the purpose of illustration only, in the following examples.

Example 1

Preparation and Screening of a Zinc Finger Phage Display Library

A powerful method of selecting DNA binding proteins is the cloning of peptides (Smith (1985) Science 228, 1315-1317), or protein domains (McCafferty *et al.*, (1990) Nature 348:552-554; Bass *et al.*, (1990) Proteins 8:309-314), as fusions to the minor coat protein (pIII) of bacteriophage fd, which leads to their expression on the tip of the capsid. A phage

display library is created comprising variants of the middle finger from the DNA binding domain of Zif268.

Materials And Methods

Construction And Cloning Of Genes. In general, procedures and materials are in accordance with guidance given in Sambrook *et al.*, Molecular Cloning. A Laboratory Manual, Cold Spring Harbor, 1989. The gene for the Zif268 fingers (residues 333-420) is assembled from 8 overlapping synthetic oligonucleotides (see Choo and Klug, (1994) PNAS (USA) 91:11163-67), giving *Sfi*I and *Not*I overhangs. The genes for fingers of the phage library are synthesised from 4 oligonucleotides by directional end to end ligation using 3 short complementary linkers, and amplified by PCR from the single strand using forward and backward primers which contain sites for *Not*I and *Sfi*I respectively. Backward PCR primers in addition introduce Met-Ala-Glu as the first three amino acids of the zinc finger peptides, and these are followed by the residues of the wild type or library fingers as required. Cloning overhangs are produced by digestion with *Sfi*I and *Not*I where necessary. Fragments are ligated to 1μg similarly prepared Fd-Tet-SN vector. This is a derivative of fd-tet-DOG1 (Hoogenboom *et al.*, (1991) Nucleic Acids Res. 19, 4133-4137) in which a section of the pelB leader and a restriction site for the enzyme *Sfi*I (underlined) have been added by site-directed mutagenesis using the oligonucleotide:

5' CTCCTGCAGTTGGACCTGTGCCATGGCCGGCTGGGCCGCATAGAATGG AACAACTAAAGC 3' (Seq ID No. 1)

which anneals in the region of the polylinker. Electrocompetent DH5α cells are transformed with recombinant vector in 200ng aliquots, grown for 1 hour in 2xTY medium with 1% glucose, and plated on TYE containing 15μg/ml tetracycline and 1% glucose.

Figure 1 shows the amino acid sequences of the three zinc fingers derived from Zif268 used in the phage display library of the present invention. The top and bottom rows represent the sequence of the first and third fingers respectively. The middle row represents the sequence of the middle finger. The randomised positions in the α-helix of the middle finger have residues marked 'X'. The amino acid positions are numbered relative to the first helical

residue (position 1). For amino acids at positions -1 to +8, excluding the conserved Leu and His, codons are equal mixtures of (G,A,C)NN: T in the first base position is omitted in order to avoid stop codons, but this has the unfortunate effect that the codons for Trp, Phe, Tyr and Cys are not represented. Position +9 is specified by the codon A(G,A)G, allowing either Arg or Lys. Residues of the hydrophobic core are circled, whereas the zinc ligands are written as white letters on black circles. The positions forming the β-sheets and the α-helix of the zinc fingers are marked below the sequence.

Phage Selection. Colonies are transferred from plates to 200ml 2xTY/Zn/Tet (2xTY containing 50μM $Zn(CH_3COO)_2$ and 15μg/ml tetracycline) and grown overnight. Phage are purified from the culture supernatant by two rounds of precipitation using 0.2 volumes of 20% PEG/2.5M NaCl containing 50μM $Zn(CH3.COO)_2$, and resuspended in zinc finger phage buffer (20mM HEPES pH7.5, 50mM NaCl, 1mM $MgCl_2$ and 50μM $Zn(CH3.COO)_2$). Streptavidin-coated paramagnetic beads (Dynal) are washed in zinc finger phage buffer and blocked for 1 hour at room temperature with the same buffer made up to 6% in fat-free dried milk (Marvel). Selection of phage is over three rounds: in the first round, beads (1 mg) are saturated with biotinylated oligonucleotide (~80nM) and then washed prior to phage binding, but in the second and third rounds 1.7nM oligonucleotide and 5μg poly dGC (Sigma) are added to the beads with the phage. Binding reactions (1.5ml) for 1 hour at 15°C are in zinc finger phage buffer made up to 2% in fat-free dried milk (Marvel) and 1% in Tween 20, and typically contained $5x10^{11}$ phage. Beads are washed 15 times with 1ml of the same buffer. Phage are eluted by shaking in 0.1M triethylamine for 5min and neutralised with an equal volume of 1M Tris pH7.4. Log phase *E. coli* TG1 in 2xTY are infected with eluted phage for 30min at 37°C and plated as described above. Phage titres are determined by plating serial dilutions of the infected bacteria.

Sequencing Of Selected Phage. Single colonies of transformants obtained after three rounds of selection as described, are grown overnight in 2xTY/Zn/Tet. Small aliquots of the cultures are stored in 15% glycerol at -20°C, to be used as an archive. Single-stranded DNA is prepared from phage in the culture supernatant and sequenced using the Sequenase™ 2.0 kit (U.S. Biochemical Corp.).

Example 2

Isolation of zinc fingers capable of C-T differentiation

The phage are selected against oligonucleotides comprising the sequences GCGGCGGCG and GCGGTGGCG. some zinc finger DNA-binding domains are selected which bound both sequences equally well (Fig. 1b, c). However, two additional zinc finger families are isolated which are capable of differential binding to the two closely related sites (Fig. 1b, c). Sequence-specific recognition requires discrimination of the central base in the binding site by amino acids in position 3 of the recognition helix of the selected zinc fingers, and it is noted that aspartate is selected to bind opposite cytosine in the triplet GCG, while alanine is selected opposite thymine in the triplet GTG. The correlation between thymine and alanine is particularly significant, as it implies a van der Waals interaction between the amino acid side-chain and the 5-methyl group of the base. Indeed, when thymine is mutated to deoxyuracil in the binding sites of such fingers there is a dramatic decrease in the strength of the intermolecular interaction (Fig 1c). This shows that these zinc fingers are capable of specifically recognising a 5-methyl group, and suggests that similar fingers might be selected which bind 5-meC by the same token.

Example 3

Selection of 5-methylcytosine-specific zinc fingers

The phage display library is screened with the synthetic binding site GCGGMGGCG, containing a 5-meC base analogue (M). After 5 rounds of selection, zinc finger phage are tested for binding to 5-meC and cytosine in the context of the above site, and those capable of specifically binding the methylated site are sequenced in the region of the zinc

finger gene. Two different clones are isolated, which are identical to the DNA-binding domains previously selected using the binding site GCGGTGGCG.

Hence the various zinc finger phage selections described above yield different fingers able to bind the generic DNA sequence GCGGNGGCG, where N is either thymine, cytosine or 5-meC. A full complement of fingers is selected for recognition of the cytosine/5-meC pair in the above context, some of which recognise one type of base exclusively, while others bound both bases equally well (Figures 1c and 2).

The zinc finger amino acid residues which are selected by the interaction between the randomised recognition helix and the central base of the DNA binding site are rationalised in terms of previously elucidated zinc finger-DNA recognition rules. Fingers with alanine in position +3 of the recognition helix specifically bind 5-meC and thymine owing to a tight hydrophobic interaction between the side chain and the 5-methyl group which is present in both bases. In contrast, a finger with valine in position +3 is also able to accommodate cytosine in addition to the two methylated bases, by the use of different rotamers. Fingers with aspartate in position +3 bind cytosine specifically, for example by forming a ring structure which packs against the pyrimidine as is observed in the refined crystal structure of Zif268.

Example 4

Selection of 5-meC Specific Zinc Fingers using Cross-Strand Specificity

1. General Procedures

Construction of overlapping finger phage display libraries

Two zinc finger DNA binding domain libraries are constructed comprising the amino acid framework of wild-type Zif268, but containing randomisations in amino acid positions of fingers 2 and 3. The first library contains randomisations at F2 residue position 6 and F3 residue positions -1, 1, 2 and 3 and recognises sequences of the form 5'-GXX-XCG-GCG-3'. The second library additionally contains variations in F2 position 3 and F3 positions 5 and 6 and recognises sequences of the form 5'-XXX-XXG-GCG-3'. The libraries are denoted collectively as LF2/3.

The genes for the two zinc finger phage display libraries are assembled from synthetic DNA oligonucleotides by directional end-to-end ligation using short complementary DNA linkers. The oligonucleotides contain selectively randomised codons, encoding all 20 amino acids or a subset thereof, in the appropriate amino acid positions of fingers 2 and 3. The constructs are amplified by PCR using primers containing *Not I* and *Sfi I* restriction sites, digested with the above endonucleases to produce cloning overhangs, and ligated into vector Fd-Tet-SN. Electrocompetent *E. coli* TG1 cells are transformed with the recombinant vector and plated onto TYE medium (1.5% (w/v) agar, 1% (w/v) Bactotryptone, 0.5% (w/v) Bacto yeast extract, 0.8% (w/v) NaCl) containing 15 mg/ml tetracycline.

Phage selections

Tetracycline resistant colonies are transferred from plates into 2xTY medium (16g/litre Bactotryptone, 10g/litre Bacto yeast extract, 5g/litre NaCl) containing 50μM $ZnCl_2$ and 15 μg/ml tetracycline, and cultured overnight at 30°C in a shaking incubator. Cleared culture supernatant containing phage particles is obtained by centrifuging at 300g for 5 minutes.

DNAs of the form 5'-tatagtG-XXXX-GGCGtgtcacagtcagtccacacacgtc-3', and their complementary strands, are chemically synthesised and annealed in 20mM Tris-HCl, pH 8, 100mM NaCl. The DNA sequences -XXXX- represent nucleotide sequences after methylation by M.*Hae*III (GGMC) or M.*Hha*I (GMGC). Since DNA is chemically synthesised, the DNA sites used in selections incorporate 5-meC (in appropriate positions on both strands) with 100% yield. Selections are also carried out on derivatives of these sites containing thymine rather than 5-meC in the appropriate positions (and with A rather than C on the complementary strand as appropriate).

One picomole of each target site is bound to streptavidin-coated tubes (Boehringer Mannheim) in 50μl PBS containing 50μM $ZnCl_2$. Bacterial culture supernatant containing phage is diluted 1:10 in selection buffer (PBS containing 50μM $ZnCl_2$, 2% (w/v) fat-free dried milk (Marvel), 1% (v/v) Tween, 20μg/ml sonicated salmon sperm

DNA), and 1ml is applied to each tube. In order to increase the selection pressure, 50 pmol soluble (unbiotinylated) competitor sites are synthesised and added to the binding mixtures: selections for phage that bind the methylated DNA contain competitors with cytosine or thymine at the appropriate positions; selections for phage that discriminate thymine instead of 5-meC in the recognition sites of the methylase enzymes contain DNA competitors with cytosine or 5-meC at the appropriate positions. After 1 hour at 20°C, the tubes are emptied and washed 20 times with PBS containing 50μM $ZnCl_2$, 2% (w/v) fat-free dried milk (Marvel) and 1% (v/v) Tween. Retained phage are eluted in 0.1ml 0.1M triethylamine and neutralised with an equal volume of 1M Tris (pH 7.4). Logarithmic-phase *E. coli* TG1 (0.5ml) are infected with eluted phage (50ml), and cultured overnight at 30°C in 2xTY medium containing 50μM $ZnCl_2$ and 15 μg/ml tetracycline, to prepare phage for subsequent rounds of selection. After 4 rounds of selection, *E. coli* TG1 infected with selected phage are plated, individual colonies are picked and used to prepare phage for ELISA assays and DNA sequencing.

ELISA to determine nucleotide discrimination.

Binding sites are synthesised as described above, including biotinylated sites where 5-meC (M) is replaced by a C or T (with appropriate bases in the complementary strand). Two-fold dilutions of DNA are added to separate wells of a streptavidin-coated microtitre plate (Boehringer Mannheim) in 50μl PBS containing 50μM $ZnCl_2$ (PBS/Zn). Phage solution (bacterial culture supernatant diluted 1:10 in PBS/Zn containing 2% (w/v) fat-free dried milk (Marvel), 1% (v/v) Tween and 20μg/ml sonicated salmon sperm DNA) are applied to each well (50μl/well). Binding is allowed to proceed for one hour at 20°C. Unbound phage are removed by washing 6 times with PBS/Zn containing 1% (v/v) Tween, then 3 times with PBS/Zn. Bound phage are detected by ELISA using horseradish peroxidase-conjugated anti-M13 IgG (Pharmacia Biotech) and the colourimetric signal quantitated using SOFTMAX 2.32 (Molecular Devices).

ELISA using an enzymatically methylated DNA binding site.

Complementary DNA oligonucleotides containing the sequences methylated by M.HaeIII and M.HhaI are chemically synthesised and annealed as described above. The

DNA is used in binding assays without exposure to the methylases, or after reaction with either or both methylase enzymes according to the manufacturer's instructions (New England Biolabs). DNA binding sites (0.5 pmol) are added to wells of a streptavidin-coated microtitre plate (Boehringer Mannheim) in 50µl PBS containing 50µM $ZnCl_2$ (PBS/Zn). The binding of various zinc finger phage clones is assayed by ELISA as described above.

DNA sequence analysis

The coding sequence of individual zinc finger clones is amplified by PCR using external primers complementary to phage sequence. These PCR products are then sequenced manually using Thermo Sequenase cycle sequencing (Amersham Life Science).

2. Experimental Results

Design of sequence-specific zinc finger proteins which bind enzymatically methylated DNA sites.

The three-finger DNA-binding domain of transcription factor Zif268 binds the DNA sequence GCGTGGGCG. Phage display libraries of this zinc finger domain have been used to elucidate aspects of the base-recognition mechanism of zinc fingers and to select fingers which bind to predetermined DNA sequences. We have constructed a set of phage display libraries in which amino acid positions from both finger 2 (F2) and finger 3 (F3) of Zif268 are simultaneously randomised in order to evaluate the effect of inter-finger synergy on the specificity of DNA binding. These libraries, hereafter denoted collectively as LF2/3, contain variants which specifically recognised DNA sequences of the form XXXXCGGCG or GXXXCGGCG, where X is any nucleotide.

The HaeIII and HhaI methyltransferases modify the internal cytosine (shown in bold lettering) of their respective DNA recognition sequences GGCC and GCGC. We therefore designed two DNA oligos, one containing the sequence GG**C**CCGGCG and the other GC**G**CCGGCG, which included the sites required for modification by the respective methylases M.HaeIII or M.HhaI (underlined). The oligos also place these

sequences in the context of binding sites that could be used to screen LF2/3 for zinc fingers that specifically recognise the modified DNA.

The two different target DNA oligonucleotides are prepared using solid phase DNA synthesis such that 5-meC is be chemically incorporated into the appropriate positions (shown in bold lettering) with 100% yield, and a biotin group is added to the 5' terminus of one DNA strand. The synthetically modified DNAs are coupled to a solid support coated with streptavidin and used in separate phage selections as described above. After four rounds of selection, individual zinc finger clones from either selection are screened by phage ELISA for binding to the methylated form of their DNA target site and discrimination against a control oligo containing the unmodified DNA. Four different zinc finger phage clones with varying specificity are selected for further study: (i) clone zfHAE(M) preferentially binds the methylated DNA incorporating the HaeIII site; (ii) clone zfHHA(M) preferentially binds the methylated DNA incorporating the HhaI site; (iii) clone zfHAE(Y) binds the DNA incorporating the HaeIII site regardless of the methylation status; and (iv) clone zfHHA(Y) binds the DNA incorporating the HhaI site regardless of the methylation status.

Table 1 shows the sequences of the oligonucleotides used for selection and of the resulting clones obtained.

Table 1

Oligonucleotide Sequences

HAE(M)	5'-tatagtG-GGMC-GGCGtgtcacagtcagtccacacacgtc-3'
HHA(M)	5'-tatagtG-GMGC-GGCGtgtcacagtcagtccacacacgtc-3'
HAE(Y)	5'-tatagtG-GGYC-GGCGtgtcacagtcagtccacacacgtc-3'
HHA(Y)	5'-tatagtG-GYGC-GGCGtgtcacagtcagtccacacacgtc-3'
HAE(T)	5'-tatagtG-GGTC-GGCGtgtcacagtcagtccacacacgtc-3'

wherein: M = 5-meC
Y = pyrimidine (C/T/M)
R = Purine (A/G)

Zinc Finger Clones

	F1							F2							F3						
	-1	1	2	3	4	5	6	-1	1	2	3	4	5	6	-1	1	2	3	4	5	6
zfHAE(M)		R	S	D	E	L	T	R		R	S	D	D	L	S	Q		R	K	H	H	R	K	E
zfHHA(M)		R	S	D	E	L	T	R		R	S	D	D	L	T	R		Y	D	G	A	R	K	R
zfHAE(Y)		R	S	D	E	L	T	R		R	S	D	D	L	T	G		H	N	R	D	R	K	R
zfHHA(Y)		R	S	D	E	L	T	R		R	S	D	H	L	S	A		T	N	S	T	R	T	K
zfHAE(T)		R	S	D	E	L	T	R		R	S	D	D	L	S	T		R	N	D	H	R	K	T

Zinc finger phage binding for each of the above clones is titrated against different amounts of methylated and unmethylated DNA oligos to derive values of the apparent dissociation constants (K_ds) for either DNA binding site (see Figures 4 and 5). The

apparent K_d of each clone for the optimally bound DNA site(s) is in the nanomolar range, similar to that of wild-type Zif268 DNA-binding domain for its preferred target site using this assay. The K_ds obtained are shown in Table 2. Clones zfHAE(M) and zfHHA(M) preferentially bind their respective DNA target sites when 5-meC is incorporated into the correct nucleotide positions, and discriminated against the unmethylated DNA sites by factors of approximately 20-fold and 5-fold respectively. The discrimination shown by zfHAE(M) in particular is good considering the simple DNA recognition mechanism of zinc fingers, and that only a single functional group per DNA molecule has been altered. Clones zfHAE(Y) and zfHHA(Y) bind their respective target sites but do not show any preference for either the modified or unmodified forms.

The four zinc finger clones isolated by phage display using synthetic 5-meC -containing DNA target sites are next tested for binding to enzymatically methylated DNA. In this assay a single DNA fragment is used that incorporates both the GGCCCGGCG and the GCGCCGGCG zinc finger binding site sequences (Figure 6a), which additionally are substrates for methylation by M.HaeIII and M.HhaI respectively. Each zinc finger clone is tested for binding to the DNA before and after DNA modification using one or both methylases. Figure 6b shows that, in contrast to zfHAE(Y) and zfHHA(Y) which both recognise the DNA regardless of the methylation status (as would be expected), zfHAE(M) and zfHHA(M) bind only after specific methylation of the DNA by the appropriate methylase enzyme. Thus enzymatic modification of cytosine to 5-meC can act as a switch that induces specific protein-DNA complex formation.

Table 2

K_ds of each clone for target and non-target oligonucleotides

Clone	Oligonucleotide	K_d
zfHAE(M)	G-GG**M**C-GGCG	2.0 +/- 0.2nM
	G-GG**C**C-GGCG	62 +/- 29nM
zfHHA(M)	G-G**M**GC-GGCG	14 +/- 3.2nM
	G-G**C**GC-GGCG	62 +/- 22nM
zfHAE(Y)	G-GG**M**C-GGCG	6.3 +/- 1.4nM
	G-GG**C**C-GGCG	2.0+/- 0.2nM
zfHHA(Y)	G-G**M**GC-GGCG	14 +/- 2.0nM
	G-G**C**GC-GGCG	11 +/- 2.4nM

Synergistic zinc finger pairs that discriminate 5-methylcytosine from thymine.

The 5-methyl group of methylcytosine and thymine is a prominent feature of the DNA major groove which contributes important intermolecular (hydrophobic) contacts in protein-DNA interactions but is stereochemically indistinguishable in the two different bases. Consequently, zinc fingers - which frequently achieve DNA recognition by 1:1 contacts between amino acids and bases - often fail to discriminate between the two closely related bases. The phage-selected clone zfHHA(M) is one such zinc finger protein which accepts both thymine and 5-meC with almost equal affinity (Figure 5). In this case it is likely that the aromatic ring of tyrosine forms equally good hydrophobic contacts with the methyl group of either base.

One way in which zinc finger proteins could distinguish 5-meC from thymine is to discriminate the complementary nucleotide in the base-pair. Zinc finger proteins such as Zif268 make base contacts predominantly to only one DNA strand - the 'antiparallel' strand - but, importantly, they can also form 'cross-strand' contacts to certain bases on the complementary, 'parallel' strand. It has been shown that these contacts can make important contributions to DNA-binding specificity. Thus the zinc fingers of Zif268 and

related proteins can be regarded as binding to overlapping 4bp subsites, where the specificity for the base-pair at the boundary between adjacent subsites potentially arises via contacts from two synergistic zinc fingers to each of the nucleotides in the base-pair (Figure 3). Therefore a zinc finger protein can distinguish a 5-meC:G base-pair from a T:A base-pair provided they are positioned at the overlap between adjacent DNA subsites, such that a contact to the 'parallel' strand can be made.

This is the case for the DNA binding site GGMCCGGCG in which the 5-meC base (bold) is discriminated from thymine by zinc finger clone zfHAE(M). According to the conventional model of zinc finger-DNA recognition, based on the crystal structure of the Zif268-DNA complex and subsequent biochemical experiments, the 5-meC base in the binding site is contacted by the glutamine residue in α-helical position +6 of finger 2 (Figure 3). Additionally, the complementary guanine can be recognised using a synergistic contact from the histidine residue in α-helical position +2 of finger 3 (Figure 3).

In order to investigate further the discrimination between 5-meC and thymine, another zinc finger clone is selected, zfHAE(T), which is specific for thymine instead of 5-meC in the context of the above binding site. This clone makes use of a cross-strand contact from aspartate in position +2 of finger 3 to recognise adenine in the 'parallel' strand. In this respect zfHAE(T) is remarkably like the wild-type Zif268 DNA-binding domain, whose zinc fingers each have an Arg-Ser-Asp triad that makes inter- and intra- molecular contacts including cross-strand contacts from the aspartate. Discrimination in favour of thymine by zfHAE(T) is relatively stronger than discrimination for 5-meC by zfHAE(M), presumably owing to the stabilising effect of intramolecular (protein-protein buttressing) interactions and the favourable geometry of this network of contacts.

The dissociation constants for the interactions seen between zfHAE(M), zfHHA(M) and zfHAE(T) and 5-meC or T oligonucleotides are set forth in Table 3.

Table 3

K_ds of each clone for 5-meC and T oligonucleotides

Clone	Oligonucleotide	K_d
zfHAE(M)	G-GG**M**C-GGCG	2.0 +/- 0.2nM
	G-GG**T**C-GGCG	27 +/- 4.4nM
zfHHA(M)	G-G**M**GC-GGCG	14 +/- 3.2nM
	G-G**T**GC-GGCG	6.1 +/- 4.5nM
zfHAE(T)	G-GG**M**C-GGCG	3.4 +/- 0.5nM
	G-GG**T**C-GGCG	n/a

Example 5

Methylcytosine-specific restriction enzyme

Phage-selected or rationally designed zinc finger domains which recognise modified bases, including 5-meC, can be converted to restriction enzymes which cleave DNA containing those modified bases, including 5-meC. This is achieved by coupling a modified base-specific zinc finger to a cleavage domain of a restriction enzyme or other nucleic acid cleaving moiety.

A method of converting zinc finger DNA-binding domains to chimaeric restriction endonucleases has been described in Kim, *et al.*, (1996) Proc. Natl. Acad. Sci. USA 93:1156-1160. In order to demonstrate the applicability of methylcytosine-specific zinc fingers to restriction enzymes, a fusion is made between the catalytic domain of Fok I as described by Kim *et al.* and the 5-meC specific zinc finger described in Example 3. Fusions of the 5-meC zinc finger nucleic acid-binding domain to the catalytic domain of Fok I restriction enzyme results in a novel endonuclease which cleaves DNA adjacent to the DNA recognition sequence of the zinc finger, namely GCGGMGGCG.

The oligonucleotides GCGGMGGCG and GCGGCGGCG are synthesised and ligated to random DNA sequences. After incubation with the zinc finger restriction enzyme, the nucleic acids are analysed by gel electrophoresis. Bands indicating cleavage of the nucleic acid at a position corresponding to the location of the oligonucleotide GCGGMGGCG are visible with the methylated, but not the unmethylated, nucleic acid.

In a further experiment, the 5-meC-specific zinc finger is fused to an amino terminal copper/nickel binding motif. Under the correct redox conditions (Nagaoka, M., *et al.*, (1994) J. Am. Chem. Soc. 116:4085-4086), sequence-specific DNA cleavage is observed, only in the presence of 5-meC containing DNA incorporating the oligonucleotide GCGGMGGCG.

Example 6

Determination of methylase activity in vivo

A reporter systems is produced which produces a reporter signal conditionally depending on the activity of a DNA methylase.

A transient transfection system using zinc finger transcription factors is produced as described in Choo, Y., *et al.*, (1997) J. Mol. Biol 273:525-532. This system comprises an expression plasmid which produces a 5-meC specific phage-selected zinc finger fused to the activation domain of HSV VP16, and a reporter plasmid which contain the recognition sequence of the zinc finger upstream of a CAT reporter gene.

Thus, a zinc finger which recognises the DNA sequence GCGGCCGCG selected by phage display as described in Choo, Y. & Klug, A. (1994) Proc. Natl. Acad. Sci. U.S.A. 91:11163-11167. By the method of the preceding examples, a further zinc finger is selected which is capable of binding to the sequence GCGGMCGCG where the central base M is 5-meC, and used to construct transcription factors as described in the foregoing.

A transient expression experiment is conducted, wherein the CAT reporter gene on the reporter plasmid is placed downstream of the sequence GCGGCCGCG. The reporter plasmid is cotransfected with a plasmid vector expressing the zinc finger-HSV fusion under the control of a constitutive promoter. No activation of CAT gene expression is observed.

However, when the same experiment is conducted in the presence of Hae III methylase, CAT expression is observed as a result of the methylation of GCGGCCGCG to form GCGGMCGCG, and consequent binding of the zinc finger transcription factor to its recognition sequence.

Claims

1. A zinc finger polypeptide which binds to a target DNA sequence containing a modified base but not to an identical sequence containing the equivalent unmodified base.

2. A polypeptide according to claim 1, wherein the target DNA sequence comprises a triplet having 5-meC at the central position, and binding to the 5-meC residue by an α-helical zinc finger binding motif in the polypeptide is achieved by placing an Ala residue at position +3 of the α-helix.

3. A method for preparing a DNA binding polypeptide of the Cys2-His2 zinc finger class capable of binding to a DNA triplet in target DNA sequence comprising 5-meC as the central residue in the target DNA triplet, wherein binding to the 5-meC residue by an α-helical zinc finger DNA binding motif of the polypeptide is achieved by placing an Ala residue at position +3 of the α-helix of the zinc finger.

4. A method for preparing a DNA binding polypeptide of the Cys2-His2 zinc finger class capable of binding to a DNA triplet in target DNA sequence comprising 5-meC, but not to an identical triplet comprising unmethylated C, wherein binding to each base of the triplet by an α-helical zinc finger DNA binding motif in the polypeptide is determined as follows:

a) if the 5' base in the triplet is G, then position +6 in the α-helix is Arg and/or position ++2 is Asp;

b) if the 5' base in the triplet is A, then position +6 in the α-helix is Gln or Glu and ++2 is not Asp;

c) if the 5' base in the triplet is T, then position +6 in the α-helix is Ser or Thr and position ++2 is Asp; or position +6 is a hydrophobic amino acid other than Ala;

d) if the 5' base in the triplet is C, then position +6 in the α-helix may be any amino acid, provided that position ++2 in the α-helix is not Asp;

e) if the central base in the triplet is G, then position +3 in the α-helix is His;

f) if the central base in the triplet is A, then position +3 in the α-helix is Asn;

g) if the central base in the triplet is T, then position +3 in the α-helix is Ala, Ser, Ile, Leu, Thr or Val; provided that if it is Ala, then one of the residues at -1 or +6 is a small residue;

h) if the central base in the triplet is 5-meC, then position +3 in the α-helix is Ala, Ser, Ile, Leu, Thr or Val; provided that if it is Ala, then one of the residues at -1 or +6 is a small residue;

i) if the 3' base in the triplet is G, then position -1 in the α-helix is Arg;

j) if the 3' base in the triplet is A, then position -1 in the α-helix is Gln and position +2 is Ala;

k) if the 3' base in the triplet is T, then position -1 in the α-helix is Asn; or position -1 is Gln and position +2 is Ser;

l) if the 3' base in the triplet is C, then position -1 in the α-helix is Asp and Position +1 is Arg.

5. A method for producing a zinc finger polypeptide capable of binding to a DNA sequence comprising a modified residue, but not to an identical sequence comprising an equivalent unmodified residue, comprising the steps of:

a) providing a DNA library encoding a repertoire of zinc finger polypeptides, the DNA members of the library being at least partially randomised at one or more of the positions encoding residues -1, 2, 3 and 6 of an α-helical zinc finger binding motif of the zinc finger polypeptides;

b) displaying the library in a selection system and screening it against a target DNA sequence comprising the modified residue;

c) isolating the DNA members of the library encoding zinc finger polypeptides capable of binding to the target sequence; and

d) optionally, verifying that the zinc finger polypeptides do not bind significantly to a DNA sequence identical to the target DNA sequence but containing the equivalent unmodified residue in place of the modified residue.

6. A method according to claim 5, wherein the nucleic acid library encodes a repertoire of zinc finger polypeptides each possessing more than one zinc fingers, the nucleic acid members of the library being at least partially randomised at one or more of the positions encoding residues -1, 2, 3 and 6 of the α-helix in a zinc finger and at one or more of the positions encoding residues -1, 2, 3 and 6 of the α-helix in a further zinc finger of the zinc finger polypeptides.

7. A method according to claim 5 or claim 6, wherein the modified residue is 5-meC and the unmodified residue is C.

8. A method according to claim 5 or claim 6, wherein the modified residue is U and the unmodified residue is T.

9. A method according to any one of claims 5 to 8, wherein the library is screened by phage display.

10. A method according to any one of claims 5 to 9, wherein the or each zinc finger has the general primary structure

(A) X^a **C** X_{2-4} **C** X_{2-3} F X^c X X X X L X X **H** X X X^b **H** - linker

-1 1 2 3 4 5 6 7 8 9

wherein X (including X^a, X^b and X^c) is any amino acid.

11. A method according to claim 10 wherein X^a is $^F/_Y$-X or P-$^F/_Y$-X.

12. A method according to claim 10 or claim 11 wherein X_{2-4} is selected from any one of: S-X, E-X, K-X, T-X, P-X and R-X.

13. A method according to any one of claims 10 to 12 wherein X^b is T or I.

14. A method according to any one of claims 10 to 13 wherein X_{2-3} is G-K-A, G-K-C, G-K-S, G-K-G, M-R-N or M-R.

15. A method according to any one of claims 10 to 14 wherein the linker is T-G-E-K or T-G-E-K-P.

16. A method according to any one of claims 10 to 15 wherein position +9 is R or K.

17. A method according to any one of claims 10 to 16 wherein positions +1, +5 and +8 are not occupied by any one of the hydrophobic amino acids, F, W or Y.

18. A method according to claim 17 wherein positions +1, +5 and +8 are occupied by the residues K, T and Q respectively.

19. A method for preparing a DNA binding polypeptide of the Cys2-His2 zinc finger class capable of binding to a DNA triplet in target DNA sequence comprising 5-meC, but not to an identical triplet comprising unmethylated C:

a) selecting a model zinc finger domain from the group consisting of naturally occurring zinc fingers and consensus zinc fingers; and

b) mutating the finger by the method of any one of claims 3 to 17.

20. A method according to claim 19, wherein the model zinc finger is a consensus zinc finger whose structure is selected from the group consisting of the consensus structure P Y K C P E C G K S F S Q K S D L V K H Q R T H T G, and the consensus structure P Y K C S E C G K A F S Q K S N L T R H Q R I H T G E K P.

21. A method according to claim 19 wherein the model zinc finger is a naturally occurring zinc finger whose structure is selected from one finger of a protein selected from the group consisting of Zif 268 (Elrod-Erickson *et al.*, (1996) Structure 4:1171-1180), GLI (Pavletich and Pabo, (1993) Science 261:1701-1707), Tramtrack (Fairall *et*

al., (1993) Nature 366:483-487) and YY1 (Houbaviy *et al.*, (1996) PNAS (USA) 93:13577-13582).

22. A method according to claim 21 wherein the model zinc finger is finger 2 of Zif 268.

23. A method according to any one of claims 3 to 22 wherein the binding protein comprises two or more zinc finger binding motifs, placed N-terminus to C-terminus.

24. A method according to claim 22, wherein the N-terminal zinc finger is preceded by a leader peptide having the sequence MAEEKP.

25. A method according to claim 23 or claim 24, wherein the DNA binding protein is constructed by recombinant DNA technology, the method comprising the steps of:

a) preparing a DNA coding sequence encoding two or more zinc finger binding preparable according to claim 23 or 24, placed N-terminus to C-terminus;
b) inserting the DNA sequence into a suitable expression vector; and
c) expressing the DNA sequence in a host organism in order to obtain the DNA binding protein.

26. A method according to one of claims 3 to 25 comprising the additional steps of subjecting the DNA binding protein to one or more rounds of randomisation and selection in order to improve the characteristics thereof.

27. A zinc finger polypeptide which binds to a target DNA sequence containing a modified base but not to an identical sequence containing the equivalent unmodified base, preparable by a method according to any one of claims 3 to 26.

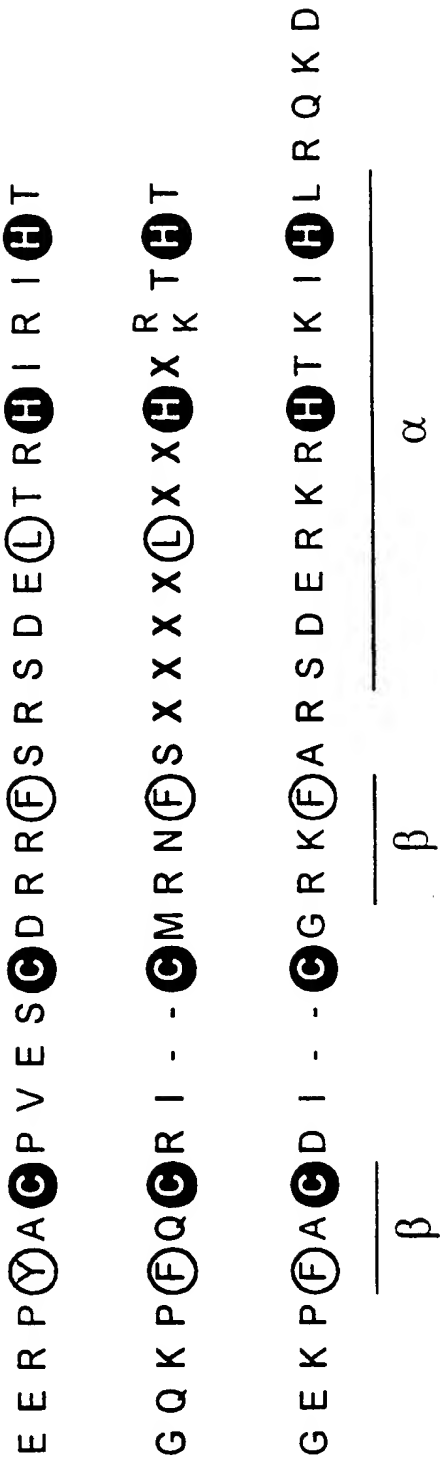

- 1 1 2 3 4 5 6 7 8 9
M A E E R P Y A C P V E S C D R R F S R S D E L T R H I R I H T
G Q K P F Q C R I - - C M R N F S X X X X L X X H X K R T H T
G E K P F A C D I - - C G R K F A R S D E R K R H T K I H L R Q K D
β
β
α

FIG. 1A

(i) G C G G M G G C G

-1 1 2 3 4 5 6 7 8 9
R A D (A) L M V H K R
R G D (A) L T S H E R

(ii) G C G G T G G C G

-1 1 2 3 4 5 6 7 8 9
R A D (A) L M V H K R
R G D (A) L T S H E R
R V D (A) L E A H R R
R E D (A) L I R H G K

(iii) G C G G C G G C G

-1 1 2 3 4 5 6 7 8 9
R G P (D) L A R H G R
R E D (V) L I R H G K

FIG. 1B

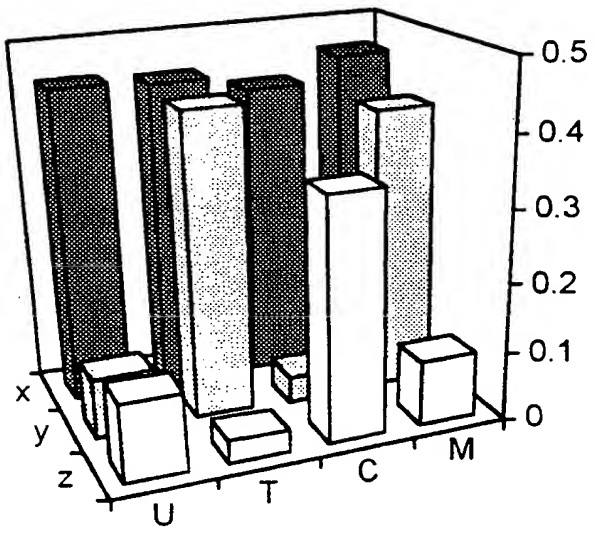


FIG. 1C

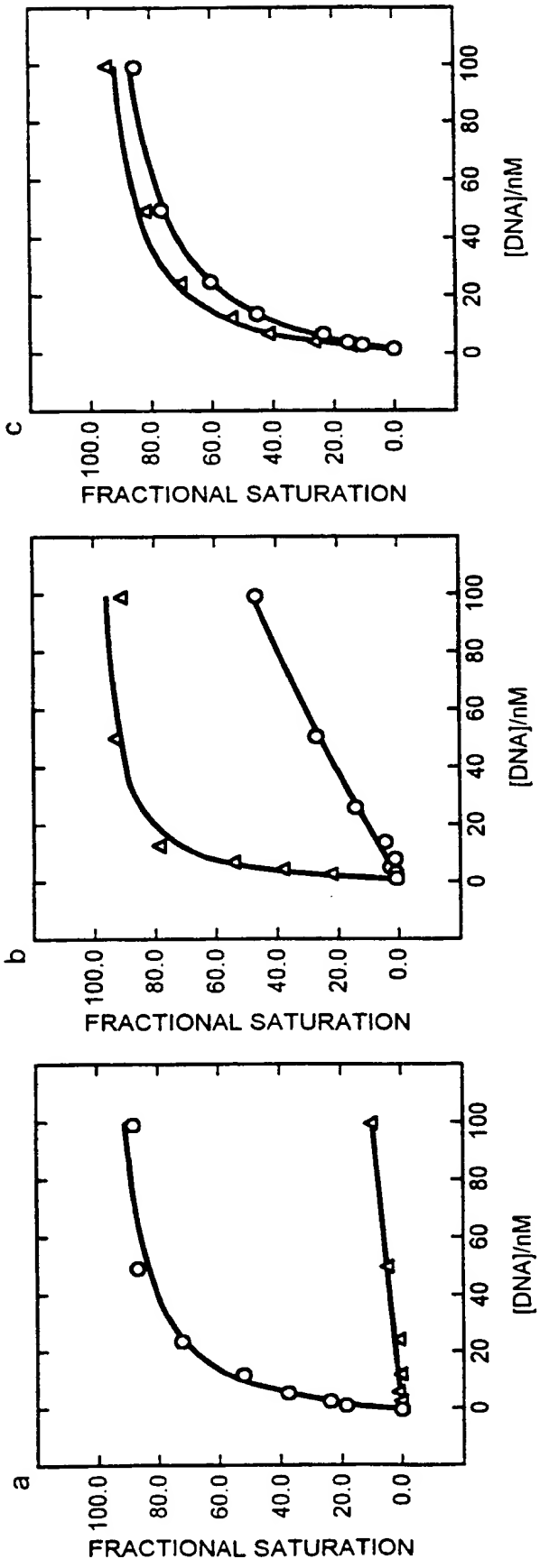
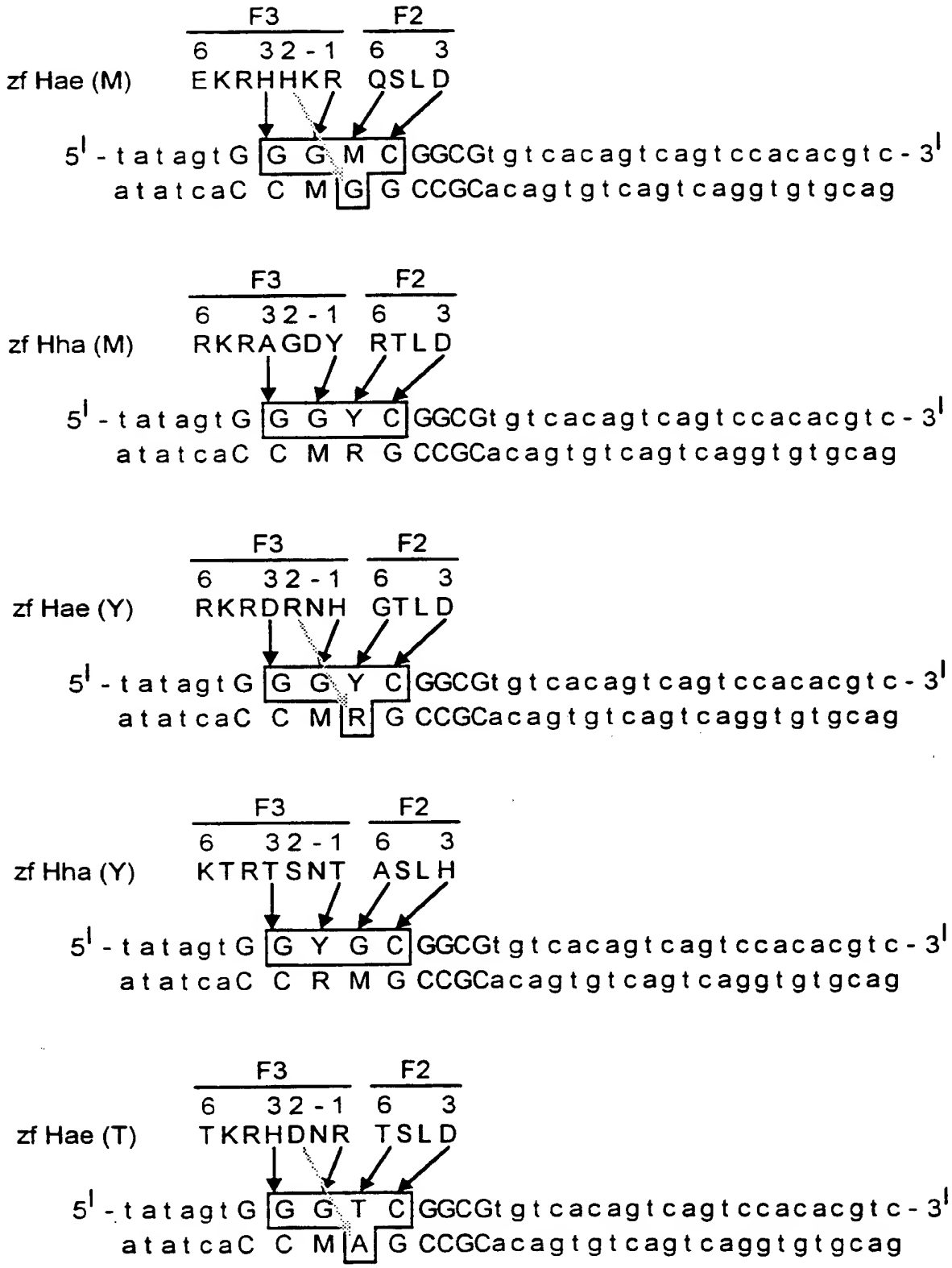

c
100.0
80.0
60.0
40.0
20.0
0.0
FRACTIONAL SATURATION
0
20
40
60
80
100
[DNA]/nM
b
100.0
80.0
60.0
40.0
20.0
0.0
FRACTIONAL SATURATION
0
20
40
60
80
100
[DNA]/nM
a
100.0
80.0
60.0
40.0
20.0
0.0
FRACTIONAL SATURATION
0
20
40
60
80
100
[DNA]/nM

FIG. 2



M=5-methylcytosine Y=pyrimidine (C/T/M) R=purine (A/G)

FIG. 3

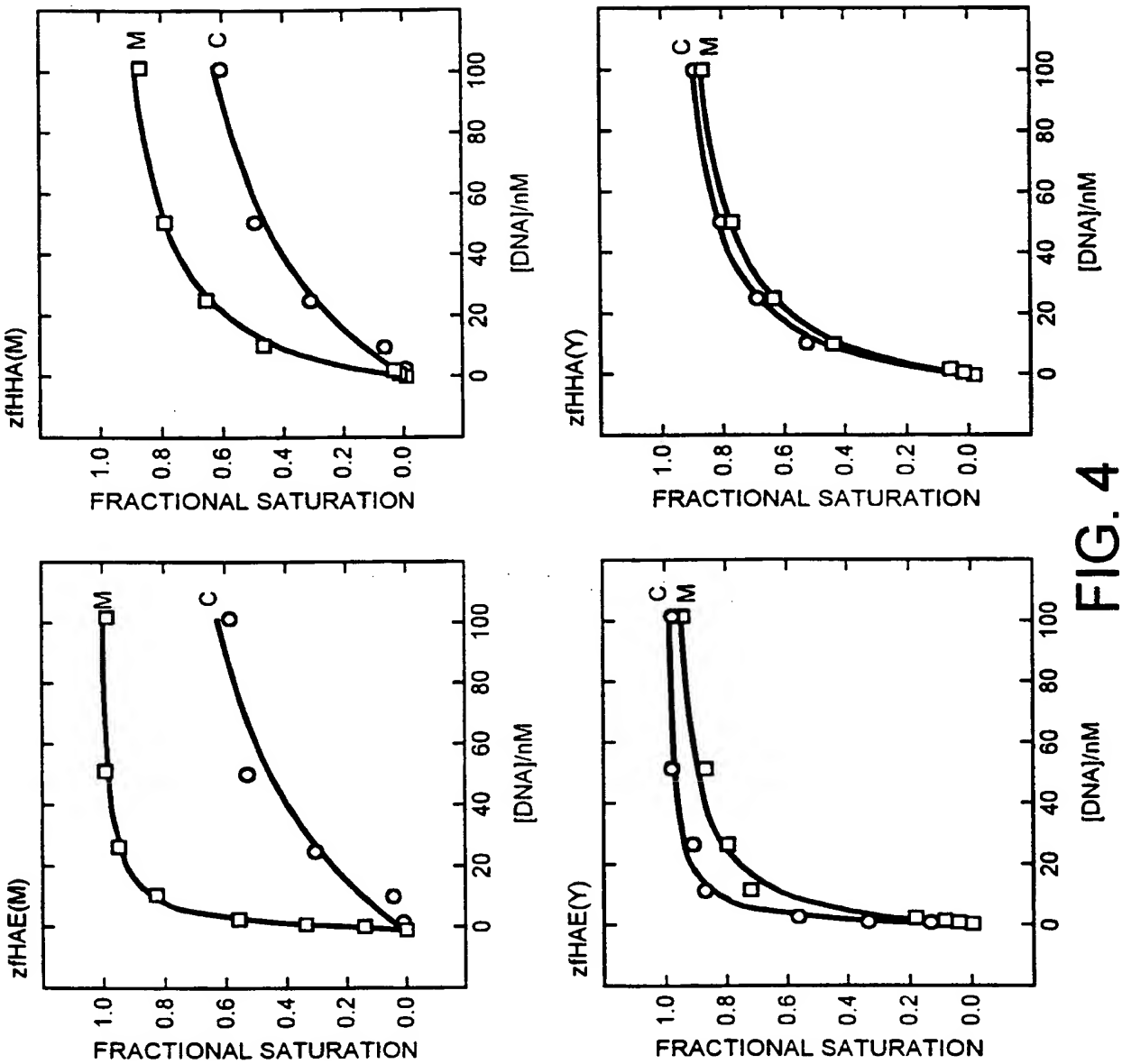

zfHHA(M)
zfHHA(Y)
zfHAE(M)
zfHAE(Y)
M
C
FRACTIONAL SATURATION
[DNA]/nM
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
100

FIG. 4

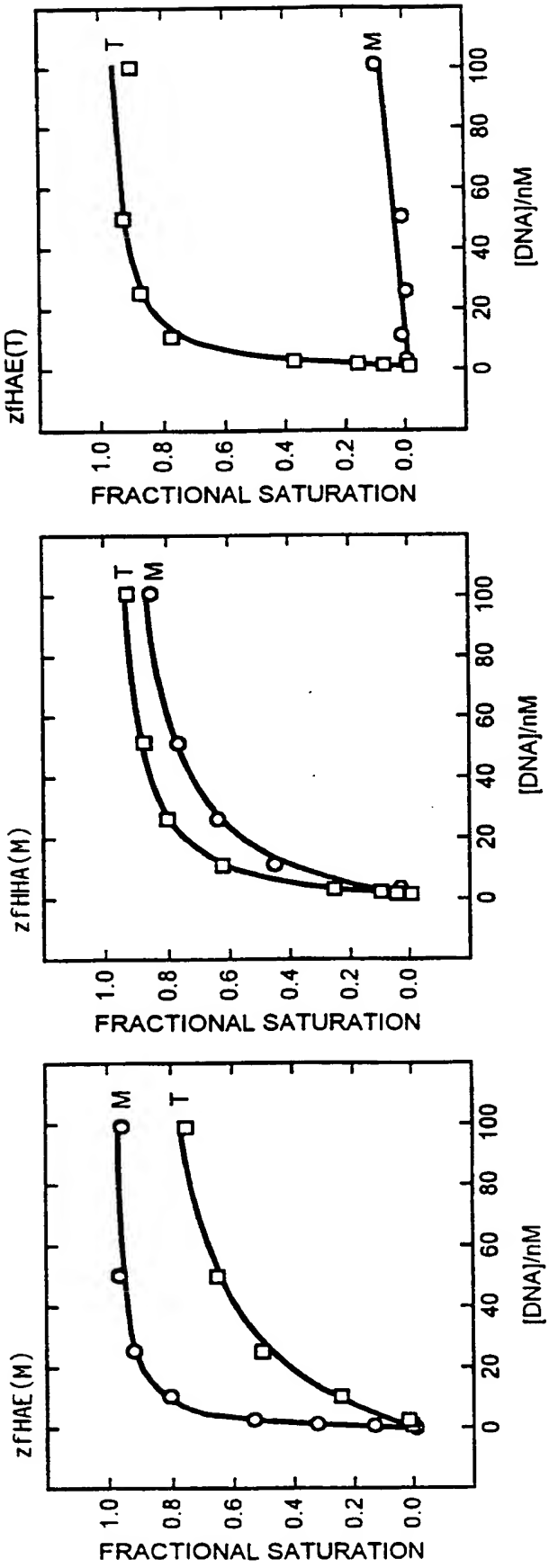

zfHAE(T)
T
M
1.0
0.8
0.6
0.4
0.2
0.0
FRACTIONAL SATURATION
0
20
40
60
80
100
[DNA]/nM
zfHHA(M)
T
M
1.0
0.8
0.6
0.4
0.2
0.0
FRACTIONAL SATURATION
0
20
40
60
80
100
[DNA]/nM
zfHAE(M)
M
T
1.0
0.8
0.6
0.4
0.2
0.0
FRACTIONAL SATURATION
0
20
40
60
80
100
[DNA]/nM

FIG. 5

THIS PAGE BLANK (USP

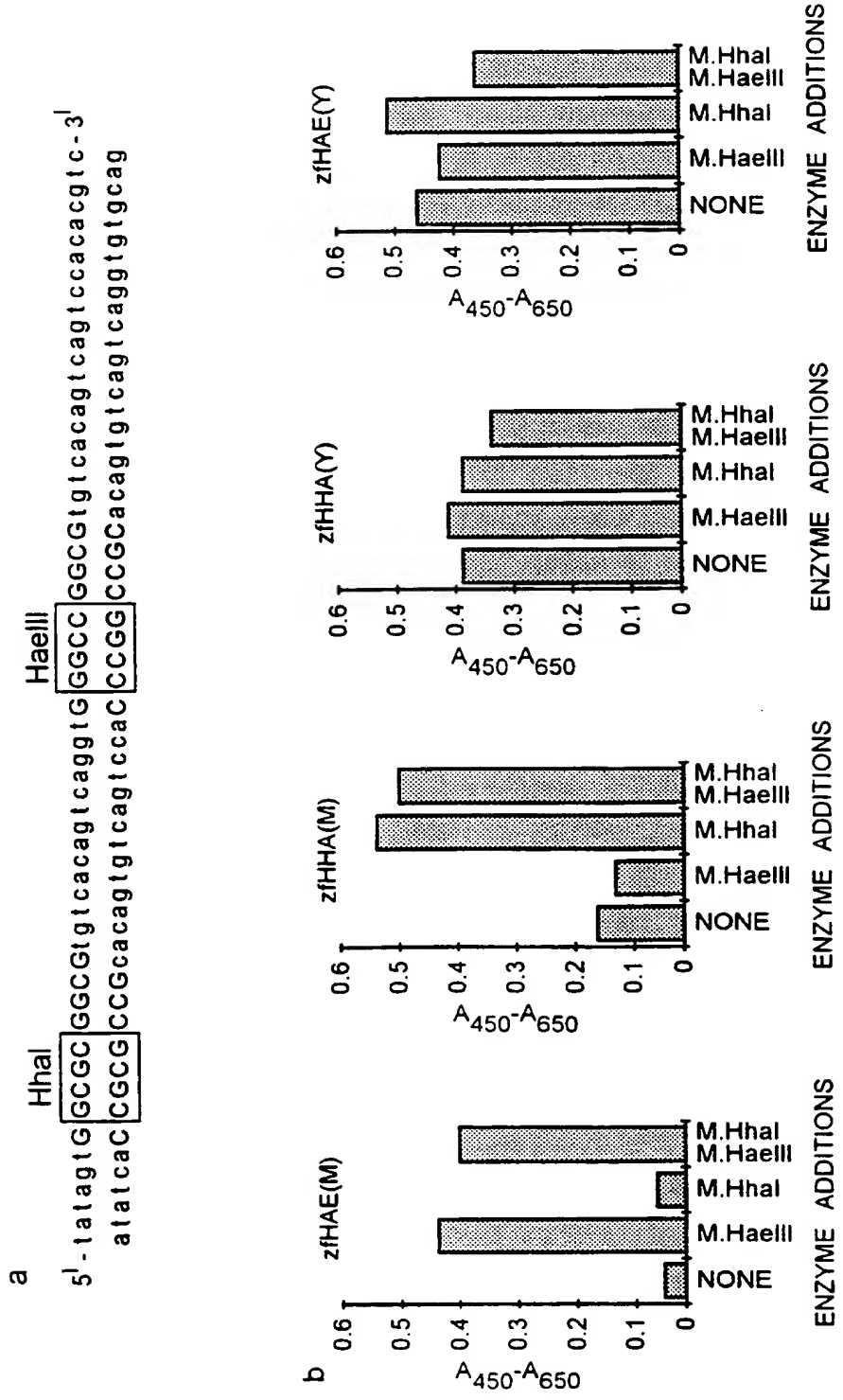

a
5'-tatagtGGCGCGGCGtgtcacagtcaggtgGGCCGGCGtgtcacagtcagtccacacgtc-3'
atatcaCCGCGCCGcacagtgtcagtccaCCGGCCGCacagtgtcagtcaggtgtgcag
HhaI
HaeIII
b
zfHAE(M)
zfHHA(M)
zfHHA(Y)
zfHAE(Y)
M.HhaI
M.HaeIII
NONE
ENZYME ADDITIONS
A450-A650
0
0.1
0.2
0.3
0.4
0.5
0.6

FIG. 6

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/12, 15/10, 15/62, 9/22, C07K 14/47, C12Q 1/68, C12Q 1/48

A3

(11) International Publication Number: **WO 99/47656**

(43) International Publication Date: 23 September 1999 (23.09.99)

(21) International Application Number: PCT/GB99/00816

(22) International Filing Date: 17 March 1999 (17.03.99)

(30) Priority Data:

9805576.7	17 March 1998 (17.03.98)	GB
9806895.0	31 March 1998 (31.03.98)	GB
9807246.5	3 April 1998 (03.04.98)	GB

(71) Applicant *(for all designated States except US)*: MEDICAL RESEARCH COUNCIL [GB/GB]; 20 Park Crescent, London W1N 4AL (GB).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: CHOO, Yen [GR/GB]; MRC Laboratory of Molecular Biology, Medical Research Council, Hills Road, Cambridge CB2 2QH (GB). ISALAN, Mark [GB/GB]; 24 Shottfield Avenue, East Sheen, London SW14 8EA (GB).

(74) Agents: MASCHIO, Antonio et al.; D. Young & Co., 21 New Fetter Lane, London EC4A 1DA (GB).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published

With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(88) Date of publication of the international search report: 25 November 1999 (25.11.99)

(54) Title: NUCLEIC ACID BINDING PROTEINS

(57) Abstract

The invention provides a method for producing a zinc finger polypeptide which binds to a target nucleic acid sequence containing a modified base but not to an identical sequence containing an equivalent unmodified base.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

INTERNATIONAL SEARCH REPORT

International Application No
PCT/GB 99/00816

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/12 C12N15/10 C12N15/62 C12N9/22 C07K14/47
C12Q1/68 C12Q1/48

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12N C07K C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	Y. CHOO: "Recognition of DNA methylation by zinc fingers" NATURE STRUCTURAL BIOLOGY, vol. 5, no. 4, April 1998 (1998-04), pages 264-265, XP002116419 WASHINGTON, DC,US; the whole document ---	1-3, 5-17,19, 21,22, 25-27
A	M ISALAN ET AL: "Synergy between adjacent zinc fingers in sequence-specific DNA recognistion" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, no. 94, 27 May 1997 (1997-05-27), page 5617 5621 XP002075337 ISSN: 0027-8424 the whole document ---	1-27

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
24 September 1999	12/10/1999

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Hornig, H

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
A	WO 96 06166 A (MEDICAL RES COUNCIL ;CHOO YEN (SG); KLUG AARON (GB); GARCIA ISIDRO) 29 February 1996 (1996-02-29) the whole document	1-27
A	Y CHOO ET AL: "Physical basis of a protein-DNA recognition code" CURRENT OPINION IN STRUCTURAL BIOLOGY, vol. 1, no. 7, 1 February 1997 (1997-02-01), page 117 125 XP002075338 ISSN: 0959-440X the whole document	1-27
A	F. RADTKE ET AL.: "Differential sensitivity of zinc finger transcription factors MTF-1, Sp1 and Krox-20 to CpG methylation of their binding sites" BIOL. CHEM. HOPPE-SEYLER, vol. 377, January 1996 (1996-01), pages 47-56, XP002116420 WALTER DE GRUYTER & CO., BERLIN, BRD the whole document	1-27
A	O. SAKATSUME ET AL.: "Binding of THZif-1, a MAZ-like zinc finger protein to the nuclease-hypersensitive element in the promoter region of the c-myc protooncogene" J. BIOL. CHEM., vol. 271, no. 49, 6 December 1996 (1996-12-06), pages 21322-31333, XP002116421 AM. SOC. BIOCHEM. MOL.BIOL.,INC.,BALTIMORE,US the whole document	1-27
A	G.-L. XU AND T.H. BESTOR: "Cytosine methylation targetted to pre-determined sequences" NATURE GENETICS, vol. 17, no. 4, December 1997 (1997-12), pages 376-378, XP002116422 NATURE PUBLISHING CO., NEW YORK, US the whole document	1-27

-/--

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	Y.-G. KIM ET AL.: "Hybrid restriction enzymes: Zinc finger fusions to Fok I cleavage domain" PROC. NATL. ACAD. SCI., vol. 93, no. 3, 6 February 1996 (1996-02-06), pages 1156-1160, XP002116423 NATL. ACAD. SCI.,WASHINGTON,DC,US; cited in the application the whole document	1-27
P,A	WO 98 53057 A (MEDICAL RES COUNCIL ;ISALAN MARK (GB); CHOO YEN (GB); KLUG AARON ( ) 26 November 1998 (1998-11-26) the whole document	
P,A	WO 98 53058 A (MEDICAL RES COUNCIL ;ISALAN MARK (GB); CHOO YEN (GB); KLUG AARON ( ) 26 November 1998 (1998-11-26) the whole document	
P,A	WO 98 53059 A (MEDICAL RES COUNCIL ;ISALAN MARK (GB); CHOO YEN (GB); KLUG AARON ( ) 26 November 1998 (1998-11-26) the whole document	
P,A	WO 98 53060 A (MEDICAL RES COUNCIL ;ISALAN MARK (GB); CHOO YEN (GB); KLUG AARON ( ) 26 November 1998 (1998-11-26) the whole document	

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/GB 99/00816

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9606166 A	29-02-1996	AU 1003799 A	22-04-1999
		AU 698152 B	22-10-1998
		AU 3229195 A	14-03-1996
		CA 2196419 A	29-02-1996
		EP 0781331 A	02-07-1997
		JP 10504461 T	06-05-1998
WO 9853057 A	26-11-1998	AU 7542298 A	11-12-1998
		AU 7542398 A	11-12-1998
		AU 7542698 A	11-12-1998
		WO 9853058 A	26-11-1998
		WO 9853060 A	26-11-1998
WO 9853058 A	26-11-1998	AU 7542298 A	11-12-1998
		AU 7542398 A	11-12-1998
		AU 7542698 A	11-12-1998
		WO 9853057 A	26-11-1998
		WO 9853060 A	26-11-1998
WO 9853059 A	26-11-1998	AU 7542498 A	11-12-1998
WO 9853060 A	26-11-1998	AU 7542298 A	11-12-1998
		AU 7542398 A	11-12-1998
		AU 7542698 A	11-12-1998
		WO 9853057 A	26-11-1998
		WO 9853058 A	26-11-1998